# X-RAY CRYSTALLOGRAPHIC ANALYSES OF THE STRUCTURES OF TWO HEME PROTEINS

**by**

**Andrew John Sutherland-Smith**

*A thesis submitted in partial satisfaction of the requirements for the degree of*

Doctor of Philosophy

*in the*

Institute of Molecular BioSciences

*at*

MASSEY UNIVERSITY, NEW ZEALAND

June, 1998

## ABSTRACT

During human development three embryonic hemoglobins are synthesised prior to formation of the placenta. These hemoglobins function to scavenge oxygen from the mother's interstitial fluid enabling embryonic respiration. The human Gower II embryonic haemoglobin ($\alpha_2\varepsilon_2$) has been crystallized in its carbonmonoxy form, and its structure determined by X-ray crystallography. The structure was solved by molecular replacement and refined at 2.9 Å. The Gower II hemoglobin tetramer is intermediate between the adult hemoglobin R and R2 states, though closer to R2. The tertiary structure of the $\alpha$ subunit is essentially identical when compared to that found in the adult ($\alpha_2\beta_2$) and fetal ($\alpha_2\gamma_2$) hemoglobins. The embryonic $\varepsilon$ subunit has a very similar structure to the homologous adult $\beta$ and fetal $\gamma$ subunits, although with small differences at the N-terminus and in the A helix. Amino acid substitutions can be identified that may play a role in the altered response of the Gower II haemoglobin to allosteric effectors, in particular chloride ions.

Nitrite reductase from *Pseudomonas stutzeri* is a periplasmic heme enzyme responsible for the reduction of nitrite to nitric oxide. This reaction is the second step in the bacterial denitrification pathway, during which nitrate acts as the terminal electron acceptor for anaerobic respiration and is consequently reduced to nitrogen gas. Nitrite reductase from *Pseudomonas stutzeri* JM300 has been crystallized in the oxidised state and X-ray diffraction data collected to a resolution of 2.8 Å. The structure has been solved by the method of molecular replacement. The structure of the enzyme is dimeric, with each monomer comprised of two domains. The smaller N-terminal domain covalently binds a *c* heme group within an all $\alpha$-helical fold similar to that of the class I *c*-type cytochromes. The larger C-terminal domain consists of an eight-bladed $\beta$-propeller structure that coordinates a $d_1$ heme, a cofactor unique to this class of enzyme. The relative positions of the two domains, and hence the orientations of the bound heme groups are markedly different compared to homologous enzymes from other species.

## ACKNOWLEDGMENTS

Rarely in scientific research is any work done in isolation, and the studies presented here are certainly no exception. I am extremely grateful to the large number of people who have assisted me with this work.

I would like to thank very much my primary supervisor Professor Ted Baker for providing much scientific, emotional and financial help throughout the course of these studies, for introducing me to scientific research and always believing that no problem was unsurmountable. I also gratefully acknowledge the contribution of ideas and advice from my secondary supervisors Dr. Geoff Jameson and Dr. Bryan Anderson.

I would like to thank a number of people for both their scientific contributions to this work, and their friendship over my time spent at Massey University: Mrs. Heather Baker, Dr. Stanley Moore, Dr. Richard Kingston and Dr. Max Paoli. I would also like to thank all other present and former members of the Department of Biochemistry and Structural Biology lab who have provided me with both technical assistance and friendly words of advice and encouragement over the years.

I would also like to thank a number of other friends I have made during my time at Massey: Tania, Ross, Catherine, Rick, Neil, Treena, Jakki, Nick, Shaun, Matt, Maria, Heather, Michelle, Amanda, as well as numerous flat-mates and team-mates.

The studies on the embryonic hemoglobins were conducted in conjunction with Dr. Tom Brittain and Dr. Oliver Hofmann from the University of Auckland. The work on the *P. stutzeri* nitrite reductase was undertaken in collaboration with Dr. Bruce Averill from the E. C. Slater Institute, University of Amsterdam. I would like to thank these people for the time and effort spent making these collaborations a success.

I gratefully acknowledge financial support from the award of a Massey University Doctoral Scholarship, and latterly Professor Ted Baker.

Thank you to my family: Richard, Mary, Jennifer and James for all their love and encouragement.

Finally, a special thanks to Rosemary for her love, understanding, patience and support.

Thank you all

## TABLE OF CONTENTS

## CHAPTER 4
## GOWER II EMBRYONIC HEMOGLOBIN: STRUCTURE AND FUNCTION



## CHAPTER 5
## *P. STUTZERI* NITRITE REDUCTASE: AN OVERVIEW

## CHAPTER 6
## *P. STUTZERI* NITRITE REDUCTASE: STRUCTURE DETERMINATION

# LIST OF FIGURES

## CHAPTER 1



## CHAPTER 2



## CHAPTER 3



## CHAPTER 4

## CHAPTER 5



## CHAPTER 6

## CHAPTER 7

## ABBREVIATIONS

**2,3BPG** 2,3-bisphosphoglycerate
**ATP** Adenosine triphosphate
**BIS-TRIS** bis(2-hydroxyethyl)imino-tris(hydroxymethyl)methane
**cDNA** Copy deoxyribonucleic acid
**CM cellulose** Carboxymethyl cellulose
**C-terminal** Carboxy-terminal
**DEAE** Diethylaminoethyl
**DMSO** Dimethylsulphoxide
**EPR** Electron paramagnetic resonance
**$FADH_2$** Reduced flavin adenine dinucleotide
**FMN** Flavin mononucleotide
**Hb** Hemoglobin
**Hb A** Human adult hemoglobin
**Hb F** Human foetal hemoglobin
**HEM** *b* heme
**HEPES** N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
**HMC** *c* heme
**HMD** $d_1$ heme
**LSQR** Least squares refinement
**MGD** Molybdopterin guanine dinucleotide
**MLR** Maximum likelihood refinement
**MME-PEG** Polyethylene glycol monomethyl ether
**MOPS** 3-[N-Morpholino]propanesulfonic acid
**MPD** 2-Methyl-2,3-pentanediol
**MR** Molecular replacement
**mRNA** Messenger ribonucleic acid
**Mw** Molecular weight
**NAc** N-terminal acetylated
**NADH** Reduced nicotinamide adenine dinucleotide
**NAR** Nitrate reductase
**NCS** Non-crystallographic symmetry
**NIR** Nitrite reductase
**NMR** Nuclear magnetic resonance
**NOR** Nitric oxide reductase
**NOS** Nitrous oxide reductase
**N-terminal** Amino-terminal
**PCR** Polymerase chain reaction
**PDB** Protein Data Bank
**PEG** Polyethylene glycol
**$P_{O2}$** Partial pressure of oxygen
**PQQ** Pyrrolo-quinoline quinone
**RBR** Rigid body refinement
**rms** Root mean square
**RT** Room temperature
**SAR** Simulated annealing refinement
**SDS-PAGE** Sodium dodecyl sulphate - polyacrylamide gel electrophoresis
**TAPS** N-tris[Hydroxymethyl]methyl-3-aminopropanesulfonic acid

**TRIS** Tris(hydroxymethyl)aminomethane
**UV** Ultraviolet
**Ala** Alanine
**Arg** Arginine
**Asn** Asparagine
**Asp** Aspartic acid
**Cys** Cysteine
**Gln** Glutamine
**Glu** Glutamic acid
**Gly** Glycine
**His** Histidine
**Ile** Isoleucine
**Leu** Leucine
**Lys** Lysine
**Met** Methionine
**Phe** Phenylalanine
**Pro** Proline
**Ser** Serine
**Thr** Threonine
**Trp** Tryptophan
**Tyr** Tyrosine
**Val** Valine

## RELATED PUBLICATION

Some of the material presented in this thesis has been accepted for publication.

Sutherland-Smith, A. J., Baker, H. M., Hofmann, O. M., Brittain T., and Baker, E. N. (1998). Crystal structure of a human embryonic haemoglobin: the carbonmonoxy form of Gower II ($\alpha_2\varepsilon_2$) haemoglobin at 2.9 Å resolution. *Journal of Molecular Biology*. In press.

*Chapter 1*

# Heme Proteins: An Introduction

## 1.1 HEME PROTEINS IN BIOLOGY

The heme prosthetic group is widely distributed amongst enzymes and other proteins in nature, with both prokaryotes and eukaryotes utilizing this cofactor for a diverse range of protein function. A heme group comprises a tetrapyrrole porphyrin macrocyclic ligand that coordinates an iron atom. The unprotonated porphyrin acts as a two electron donor to the $Fe^{2+}$ or $Fe^{3+}$. The porphyrin ligand is not specific for iron and can also bind other metals, e.g. the magnesium tetrapyrrole cofactors found in the chlorophyll proteins.

The heme group conveys structural, electronic and spectral properties to the heme protein that cannot be achieved by polypeptide alone. The delocalized electron system of the heme allows it to function as a redox centre, and also conveys colour to the molecule to which it is bound. The properties of the heme are 'tailored', both by substituents on the porphyrin and by the nature of the one or two axial iron ligands provided by the protein. The electrostatic and hydropathic nature of the environment in which the heme is located inside the protein can also result in changes to its functional properties. The polypeptide can also modify chemical reactions at the heme by sterically affecting the manner in which the substrate can approach and bind to the iron. This flexibility of cofactor properties has resulted in heme proteins being utilised in a variety of diverse roles throughout biology. For comprehensive reviews of the types of heme proteins, their properties, and the roles they play in biology see: the PROMISE database (http://bioinf.leeds.ac.uk/promise/HAEMMAIN.html, Degtyarenko et al., 1998), or the collection edited by Eichhorn and Marzilli (Poulos, 1988; Cusanovich et al., 1988; Brunori et al., 1988; Rifkind, 1988).

Heme proteins can be divided into three broad classes based on function:

i) Oxygen transport and storage
ii) Electron transfer
iii) Enzymes

### *1.1.1 Oxygen transport and storage*

Heme proteins are utilised as vehicles for the movement and storage of oxygen in a wide range of organisms. The ability of the heme prosthetic group, when modulated by the protein environment, to reversibly bind oxygen allows these molecules to function in this role. This group of heme proteins is called the globins. The members of this group share a highly conserved protein fold (see Figure 2.1) consisting of eight helices, with the exception of the human hemoglobin $\alpha$ subunit which has seven. Vertebrates utilize myoglobin for the storage of oxygen and hemoglobin for oxygen transport. For a full discussion of human hemoglobin structure and function see Chapter 2. The globins are also found in plants, bacteria, and some fungal species. All the globins contain *b* heme as cofactor, with the exception of the chlorocruorins from the polychaete worms, which bind a chlorocruorohaem. The difference between the two is the substitution of one of the vinylic groups for a formyl side-chain (Lemberg and Falk, 1951). The *b* heme and chlorocruorohaem groups are shown in Figure 1.1.

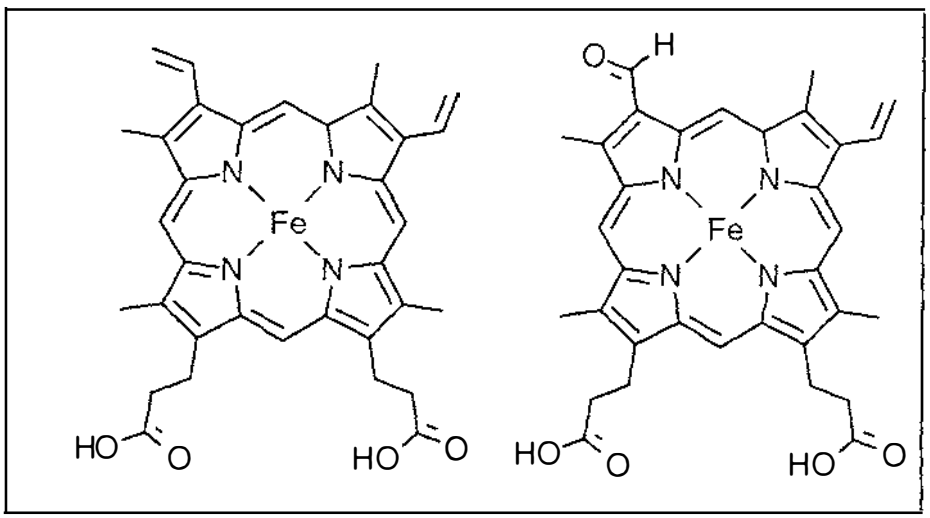



**Figure 1.1** Heme groups found in the globins: *b* heme (left) and chlorocruorohaem (right)

The globins also all share a similar mode of heme coordination, with the iron bound by an axial histidine ligand. This leaves one free coordination site on the iron, which is normally in the $Fe^{2+}$ oxidation state, to bind oxygen.

### *1.1.2 Electron transfer*

The cytochromes are a family of heme proteins that play a role in electron transport and are ubiquitously spread throughout the living world. The most well characterised of these proteins are the *c*-type cytochromes, distinguishable by their unique heme group (Figure 1.2). The properties of this protein family have been reviewed in detail (Meyer and Kamen, 1982; Pettigrew and Moore, 1987; Moore and Pettigrew, 1990). The *c* heme is covalently attached to the polypeptide chain via two thioether bonds, provided by two cysteine residues of the protein. These bonds are in addition to the protein ligands that provide the two axial ligands to the iron. The presence of a *c* heme in a protein can be identified from the polypeptide sequence by the presence of the conserved *c* heme binding motif, Cys-X-X-Cys-His. The histidine side-chain contained in this motif serves as an axial ligand to the heme iron. This heme attachment site can vary in its location in the protein sequence, and the *c* cytochromes vary in the nature of their sixth ligand, either another histidine, or a methionine.

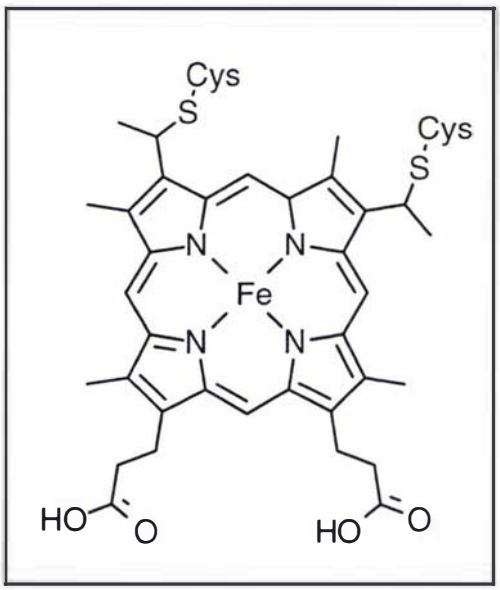



**Figure 1.2** The *c* heme group found in the *c*-type cytochromes

The main function of these proteins is to shuttle electrons via the reaction

$$Fe^{2+} \leftrightarrow Fe^{3+} + e^-$$

The electron is transferred to other proteins, some of which are membrane bound, for use in redox reactions.

The *b*-type cytochromes contain non-covalently bound *b* heme as the redox centre. This family of cytochromes is also widely spread throughout the living world and

functions in electron transfer in animals, plants and fungi. The *b* cytochromes, like the *c*-type protein, can exist as small proteins, or are found as domains in larger molecules. Consequently examples are found of both soluble and membrane-bound *b*-type cytochromes (Xia and Mathews, 1990). The *b* cytochromes have a histidine as one axial ligand and histidine or methionine completing the coordination of the iron (Mathews et al., 1979). The exceptions to these generalizations are bacterioferritin (cytochrome $b_1$) which exhibits bis-methionine coordination (Frolow et al., 1994) and cytochrome *f* in which the N-terminal $\alpha$-amino group is the sixth ligand (Martinez et al., 1994).

### *1.1.3 Enzymatic catalysis*

Heme-containing enzymes are essential components of electron transport and oxidative metabolic pathways. The reactions catalyzed include the oxygenation and hydroxylation of organic metabolites, as well as the removal of toxic peroxides. The heme groups of these enzymes generally have more negative redox potentials in comparison to the *c* cytochromes and globins. This is achieved by burying the heme inside the protein in a slightly more polar environment and utilizing electron donating proximal ligands. In contrast, the heme groups of the globins and electron transport cytochromes are generally solvent exposed. The enzymes generally provide just one amino acid ligand to the heme, leaving a coordination site vacant for the substrate to bind.

The catalases and peroxidases are *b* heme-containing enzymes. Catalase removes toxic peroxide by-products of aerobic respiration from the cell, with the heme group coordinated via a tyrosine side-chain (Vainshtein et al., 1986). The peroxidases utilize $H_2O_2$ as an electron acceptor to catalyze a variety of oxidative reactions. The peroxidases coordinate the heme with a histidine residue, with the exception of chloroperoxidase which has a cysteine as a heme ligand (Dawson, 1988).

The cytochrome $P_{450}$ class of enzymes catalyzes the incorporation of molecular oxygen into organic substrates. This often has the effect of solubilizing the product and allowing it to be removed from the cell. These enzymes contain a *b* heme coordinated by a cysteine ligand (Ravichandran et al., 1993). Nitric oxide synthase shares this method of heme coordination (Crane et al., 1998).

Cytochrome *c* oxidase is the membrane-bound terminal electron acceptor of the mitochondrial and bacterial respiratory chain. This enzyme is comprised of up to 13

subunits and contains two heme groups, *a* and $a_3$, as well as copper redox centres. The complex catalyzes the four-electron reduction of oxygen to water. This reaction results in the generation of a transmembrane proton gradient which is utilised by ATP synthase to generate ATP. The *a* heme group differs from *b* heme by the substitution of one of the vinyl groups for a much larger side-chain (Figure 1.3). The *a* heme is coordinated by two histidine residues, while the $a_3$ heme has one histidine axial ligand and a vacant site for oxygen binding (Tsukihara et al., 1996; Ostermeier et al., 1997).

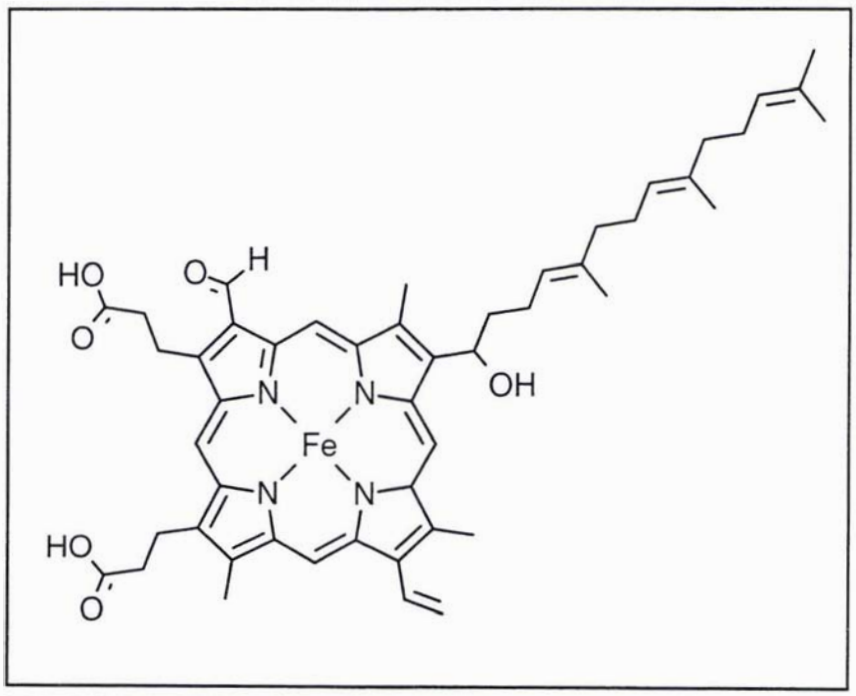



**Figure 1.3** The *a* heme group found in cytochrome *c* oxidase

There are two classes of bacterial heme enzymes that have *d*-type heme groups. The terminal oxidase complex and catalase that contain *d* heme have been isolated from many bacteria, (references summarized in Degtyarenko et al., 1998) whilst the cytochrome $cd_1$-nitrite reductase contains $d_1$ heme. (Figure 1.4). For a full discussion of the functional and physico-chemical properties of this latter enzyme see Chapter 5.

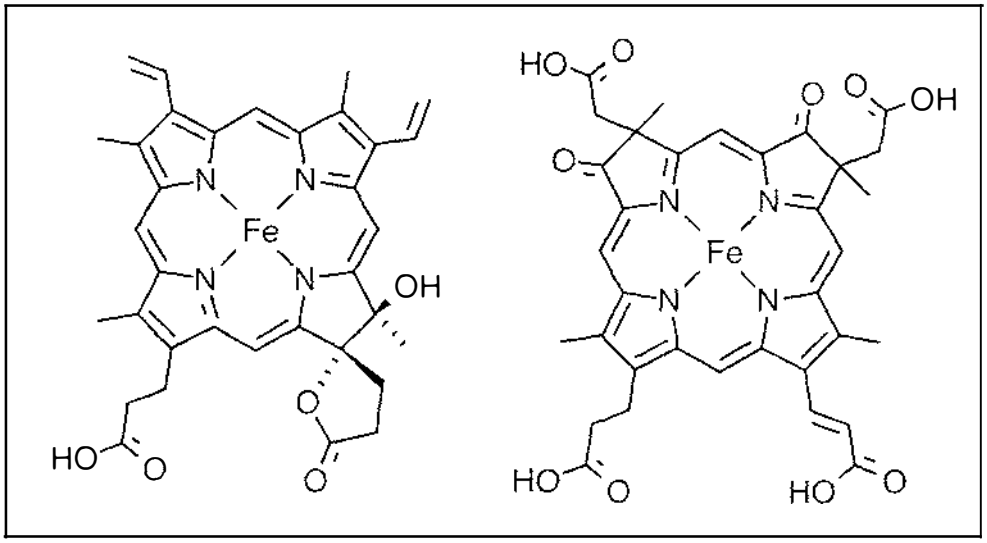



**Figure 1.4** The bacterial *d* heme groups: *d* heme (left), and $d_1$ heme (right)

The two proteins that are the subjects of this study are examples of the three broad functional types of heme proteins described above. The Gower II embryonic hemoglobin is a member of the human hemoglobin family responsible for oxygen transport in the developing embryo. The cytochrome $cd_1$ nitrite reductase from *Pseudomonas stutzeri* is an enzyme responsible for the reduction of nitrite to nitric oxide in the periplasm of the bacterial cell, as part of the denitrification pathway. This enzyme is comprised of two domains, each of which contains a heme group. One domain binds a *c* heme and is responsible for the transfer of an electron from an external electron donor to the enzyme's active site, which is a $d_1$ heme bound to the second domain where nitrite reduction occurs.

*Chapter 2*

# Hemoglobin: An Overview

## 2.1 THE IMPORTANCE OF OXYGEN IN BIOLOGY

Oxygen is essential for the majority of life-forms that inhabit our world. An aerobic metabolism is favoured, as it enables a more efficient use of available energy sources. In the presence of oxygen a far greater quantity of energy can be obtained by the oxidation of energy-rich molecules as opposed to an anaerobic metabolism.

Fatty-acid oxidation, glycolysis and the tricarboxylic acid cycle produce NADH and $FADH_2$. These two cofactors are energy-rich as they both have an electron pair with a high transfer potential. ATP is generated when these electrons are donated to molecular oxygen via a series of electron carriers in a process called oxidative phosphorylation. The oxidation of glucose to $CO_2$ and $H_2O$ generates 30 molecules of ATP, of which 26 are formed via the oxidative-phosphorylation pathway (Stryer, 1995).

There are two principal mechanisms employed by vertebrates to ensure that their cells are supplied with sufficient oxygen: a circulation system that delivers oxygen to the tissues, and the use of molecules to transport oxygen from the lungs to the cells. This overcomes the inherent problem of oxygen having a low solubility in aqueous solution.

In vertebrates the oxygen-carrying molecules are the heme proteins, myoglobin and hemoglobin. Myoglobin is found within the muscle tissue, where it serves as a reserve supply of oxygen for the respiring cells and aids the movement of oxygen within the muscle tissue. Hemoglobin is found within the red blood cells. This protein binds oxygen in the lungs and transports it to the tissues via the blood stream. Hemoglobin is also involved in the transport of carbon dioxide and hydrogen ions.

The binding of oxygen to myoglobin and hemoglobin is due to the presence of a prosthetic heme group. In the protein environment the ferrous ion is able to reversibly bind an oxygen molecule, instead of being irreversibly oxidised.

## 2.2 HEMOGLOBIN: HISTORICAL ASPECTS.

Hemoglobin has been one of the most comprehensively studied proteins over the last century. There are many books in which much of this investigation has been reviewed (Antonini et al., 1981; Bunn et al., 1977; Dickerson and Geiss, 1983; Imai, 1982; Lehmann and Huntsman, 1974; Maclean, 1978; Weatherall, 1976).

The iron content of hemoglobin had been calculated to be around 3.5% of the total weight of the protein in as early as 1925 (Englehardt, 1925). This result lead to a calculated molecular weight of the molecule of approximately 16 000 Daltons. The heme prosthetic group was isolated in 1852 as hemin chloride (Teichman, 1853) and its structure subsequently determined in 1912 (Kuster, 1912). However investigation of ligand binding to heme groups accelerated with the first heme synthesis (Fischer and Zeile, 1929) which opened up the study of synthetic model compounds.

The absorption spectrum of oxygenated hemoglobin in the red blood cells was first measured by Hoppe-Seyler in 1864 (Hoppe-Seyler, 1864) who also christened the protein 'hemoglobin' (Hb). Around the same time Stokes observed that there was a shift in the spectrum when hemoglobin was deoxygenated (Stokes, 1864). Quantitative oxygen binding measurements were conducted during the early 1900s by Bohr and Krogh (Bohr, 1904; Bohr et al., 1904), and also Barcroft and Haldane (Barcroft, 1928).

Osmotic pressure measurements (Adair, 1925) and analytical ultracentrifugation measurements (Svedberg and Fahraeus, 1926) gave rough molecular weight values of 64 to 68 kDa, suggesting that the functional unit of hemoglobin consisted of a tetramer. Technical advances enabled the detailed investigation of the spectral properties of hemoglobin, allowing the absorption spectral shifts on ligand binding to be quantified (Heilmeyer, 1943; Drabkin, 1949). These results agreed with the gasometric measurements done previously.

From the calculated molecular mass for the hemoglobin tetramer there were expected to be approximately 580 amino acids. Analysis of peptide fragments resulting from trypsin digests of hemoglobin indicated that the hemoglobin tetramer was comprised of two pairs of differing polypeptide chains (Ingram, 1955; Rhinesmith et al., 1957; Braunitzer, 1958). These were designated alpha ($\alpha$) and beta

(β), giving the hemoglobin tetramer the subunit structure $\alpha_2\beta_2$.

The primary structure was elucidated over the next few years by the technique of peptide fingerprinting after enzymatic digestion of the polypeptide chain (Braunitzer, 1958; Chernoff, 1961; Clegg et al., 1966). The α chain comprises 141 amino acids whilst the β chain is a polypeptide of 146 residues. Each chain has a heme group bound.

The crystallisation of mammalian hemoglobins had been well catalogued at the turn of the century (Reichert and Brown, 1909). It had been demonstrated that horse deoxy and oxyhemoglobin crystallised in different space groups, indicating the possibility of a conformational change in the protein molecule on ligand binding and release. Further evidence for this idea was presented when it was shown that horse deoxyhemoglobin crystals cracked when exposed to oxygen (Haurowitz, 1938), in direct contrast to crystals of myoglobin which remain unchanged.

## 2.3 THE STRUCTURE OF HEMOGLOBIN

In 1960 Perutz and his group determined the tertiary structure of horse methemoglobin at 5.5 Å resolution. This was achieved using isomorphous replacement with heavy atoms to determine the phase angles of the diffracted X-rays to calculate the Fourier synthesis (Perutz et al., 1960). The most striking feature of the subunits comprising the tetramer was that they had a very similar structure to that of the myoglobin monomer. Sperm whale myoglobin had been the first protein structure solved two years previously (Kendrew et al., 1958). The hemoglobin tetramer was 50 x 55 x 64 Å in size, arranged with a single two-fold axis down a central cavity between the $\alpha_1\beta_1$ and $\alpha_2\beta_2$ dimers. There are two pseudo-two-fold axes orthogonal to the true two-fold which rotate α subunits on to β and vice versa.

The β subunits were found to be comprised of eight helical segments (labelled A-H), (Figure 2.1). The shorter α subunits contained seven helical elements, with the small D helix not present (labelled A-H, with D omitted). The four heme pockets were all exposed at the surface.

Further work led to the determination of the structure of human deoxyhemoglobin at a similar resolution in Perutz's lab a few years later (Muirhead and Perutz, 1963). The arrangement of the subunits differed from that of methemoglobin, indicating

that there was a large conformational change upon ligand binding.

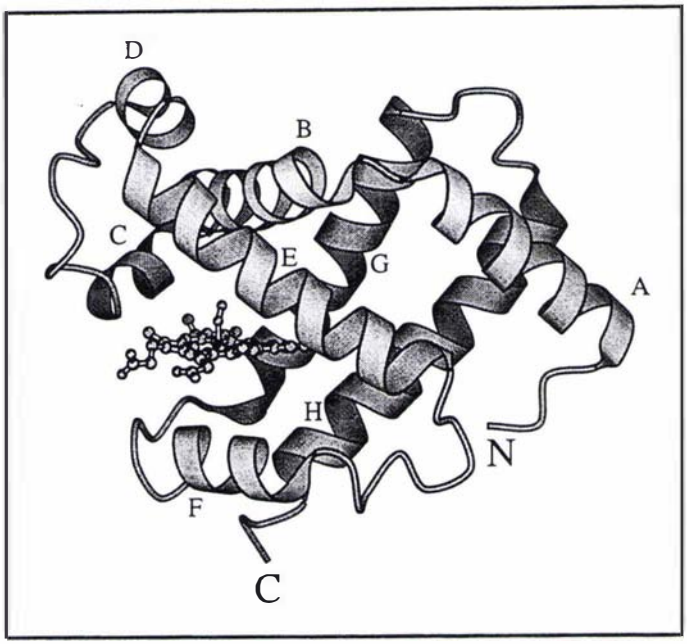



**Figure 2.1** Ribbon diagram of the human adult hemoglobin β subunit

The helices are labelled A-H. The heme group and oxygen ligand are shown in a ball and stick representation. The figure was prepared with MOLSCRIPT (Kraulis, 1991).

The liganded conformation has subsequently been termed the 'relaxed' or R state, whilst the deoxy form is the 'tense' or T state. During the oxy ↔ deoxy transition each $\alpha_1\beta_1$ and $\alpha_2\beta_2$ dimer moves as a unit, with the two halves of the molecule sliding over each other and rotating by 15°. The carboxyl termini at the end of the H helices come 7 Å closer in the liganded form, resulting in a partial closure of the central cavity of the tetramer.

Since the pioneering work of Perutz and his lab many more detailed hemoglobin structures have been solved and refined at higher resolution. For a full list of hemoglobin structures that have been determined to date see the PROMISE database (Degtyarenko et al., 1998, http://bioinf.leeds.ac.uk/promise/GLOBINS.html#PDB).

The heme is located within a pocket between the E and F helices and its iron atom is tightly bound to the Nε of histidine F8 (α87 and β92), also known as the proximal histidine (for a detailed description of adult hemoglobin see the references given at

the beginning of this chapter). The heme pocket does not have a hole large enough for oxygen to get through, but it is thought that the distal E7 (α58 and β63) histidine may shift to allow it through in some intermediate during the T↔R transition.

### *2.3.1 Changes in the stereochemistry of the heme on oxygenation*

In deoxyhemoglobin the porphyrin ring is domed, with the ferrous ion shifted ~0.6 Å out of the plane of the four nitrogen atoms towards the proximal histidine. When an oxygen molecule binds to the iron, the porphyrin flattens out with a consequent decrease in the iron to porphyrin-nitrogen bond distance. This in turn moves the metal ion into the plane of the porphyrin. This movement allows the proximal histidine to move 0.5-0.6 Å closer to the heme plane in oxyhemoglobin.

An out-of-plane Fe atom in a domed porphyrin ring is in an unfavourable conformation to bind a sixth ligand on the opposite side of the heme. For a ligand molecule to bind, the proximal histidine (F8) must straighten and move its side-chain closer to the heme. In the deoxy state it is prevented from doing this due to steric hindrance between its ε hydrogen and the heme, and also it cannot alter its 8° tilt due to its attachment to the F helix and the packing of the surrounding side-chains.

### *2.3.2 Changes on oxygenation in the tertiary conformation of the globins.*

The subunits of hemoglobin interact with each other by relatively weak noncovalent interactions such as van der Waals forces and hydrogen bonds.

#### *2.3.2.1 The α subunits*

The largest movement in going from deoxy to oxyhemoglobin occurs in the F helix in which the proximal histidine (F8) is located. Upon oxygen binding this helix moves 1.4 Å along its axis parallel to the heme plane and 0.4 Å closer to the heme, reorienting the histidine imidazole ring and aligning it with the heme normal. This movement is dampened out along the E and G helices, with the remainder of the α subunit essentially unchanged.

#### *2.3.2.2 The β subunits*

The structural changes in the β subunit are similar to those of α. The F helix moves 0.9 Å parallel to the heme and 0.4 Å closer. The heme group itself moves deeper into the heme pocket by 1.0 Å and twists slightly, moving the ligand binding site away from the Val E11 side-chain which blocks access to the iron in the deoxy con-

formation. The pattern of hydrogen bonds formed by the carboxyl terminus in the deoxy T state is disrupted and the nearby reactive Cys 93 has a different side-chain conformation. The rest of the chain remains undisturbed.

### *2.3.3 Proposed mechanism for the T to R state transition of hemoglobin*

Many papers have been written by Perutz and others with possible explanations of how ligand binding triggers the large conformational change observed for hemoglobin (Perutz, 1970; Perutz, 1972; Baldwin and Chothia, 1979; Perutz, 1979; Perutz et al., 1987). Deoxyhemoglobin is held in the T state configuration by a network of hydrogen bonds and salt bridges at the amino and carboxyl termini. With no ligand present the four heme groups are in a strain-free conformation with the histidine F8 imidazole tilted off axis and a domed heme containing an out of plane Fe.

On ligand binding to the T state, strain is introduced, which is relieved by rotating the histidine and shifting the F helix. This helix movement introduces unfavourable contacts into the globin tertiary structure which are relieved (after two or three ligands have bound) by breaking the terminal hydrogen bonds and so favouring the subunit rotation that is the T to R transition. The conformational change relieves strain on the hemes with ligand bound, but it also has the effect of placing any unliganded hemes in the tetramer in a conformation now suited to binding a sixth ligand, hence increasing their affinity for oxygen. Conversely, removing one or more of the oxygen ligands from the R state hemoglobin tetramer does not necessarily introduce any new strain, but lowers the energy barrier against a shift to the T state and reformation of the stabilising C-terminal hydrogen bonds. When the stability that reformation of these hydrogen bonds would produce becomes greater than the energy needed to force the remaining liganded hemes into the T state (tilted His) conformation, the R to T transition occurs. The strain introduced by the R to T transition can be relieved by ejecting the remaining $O_2$ molecules, so the oxygen affinity of hemoglobin is decreased.

This transformation does not appear to be as simple as the above mechanism would suggest however. Recently structural evidence for a third quaternary state of hemoglobin (R2 or Y) has been reported (Smith et al., 1991; Silva et al., 1992; Janin and Wodak, 1993; Smith and Simmons, 1994). This quaternary structure is stabilised by the mutation β99 Asp→Tyr (Hb Ypsilanti), which is a mutation in the intersubunit region of the molecule, and/or by solutions of low salt and low pH. There must be intermediate structures between the T and R states, but debate still

continues as to whether the R2 state is an intermediate on this pathway, or whether the mechanism is more likely to be T↔R↔R2 (Silva et al., 1992; Janin and Wodak, 1993; Srinivasan and Rose, 1994; Schumacher et al., 1997).

## 2.4 FUNCTIONAL PROPERTIES OF HEMOGLOBIN

### *2.4.1 Oxygen binding*

Despite all hemoglobins having a chemically and physically identical heme group and proximal histidine ligand, the oxygen binding affinity within this protein family varies greatly. For hemoglobin to function effectively it must bind oxygen with an appropriate affinity. If it does not bind tightly enough the blood will not become sufficiently oxygenated as it passes through the lungs, however if the binding is too tight then insufficient oxygen will be unloaded to the respiring tissue.

The partial pressure of oxygen in the lungs is 95 mmHg. When the blood passes through this oxygen-rich environment, the hemoglobin becomes 97% saturated (Figure 2.2). Normal blood with a hemoglobin concentration of approximately 15g/100ml would carry 20ml of oxygen per 100ml of blood. This is in comparison with the 0.3ml of oxygen which can dissolve in 100ml of aqueous solution. This arterial blood is then circulated to the capillaries which have a partial oxygen pressure of 40 mmHg. At this oxygen concentration hemoglobin is 75% saturated, hence hemoglobin releases 4.5 ml of oxygen per 100ml of blood as it circulates through the organism (Bunn et al., 1977).

The oxygen affinity is quantified as the $P_{50}$ value, which is defined as the partial pressure of oxygen at which the hemoglobin is half saturated. The greater the affinity of the hemoglobin for oxygen the lower the $P_{50}$ and vice versa.

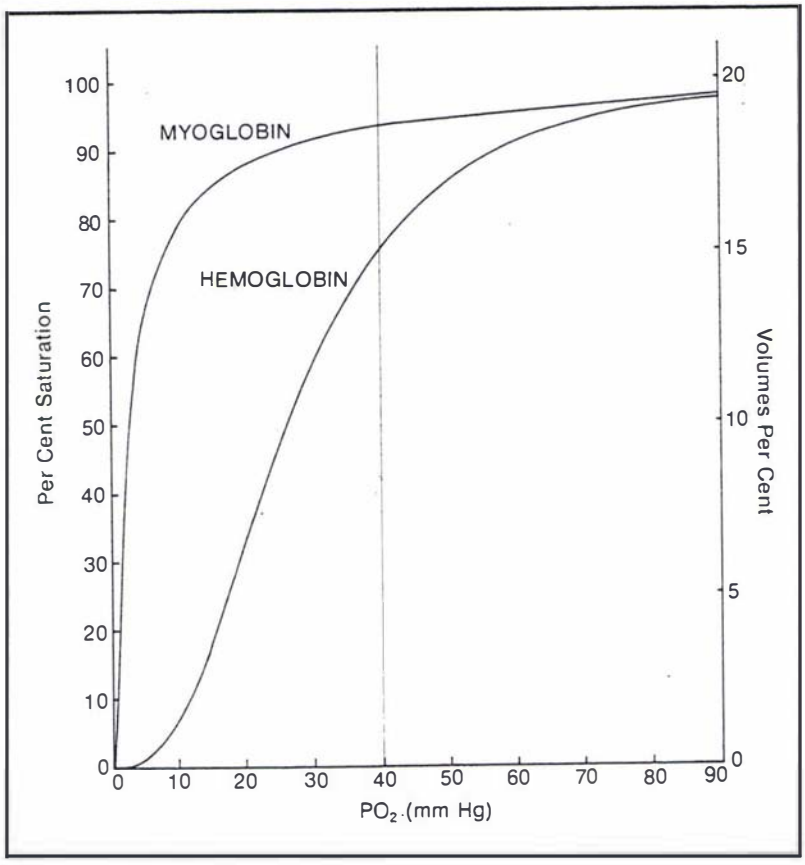



**Figure 2.2** Oxygen binding curves of hemoglobin and myoglobin

Figure reproduced from Bunn et al., 1977

This efficient mode of oxygen transport depends on the cooperativity of the hemoglobin subunits, where the binding of oxygen to one heme group results in the increased oxygen affinity of the other subunits. This is manifested in the sigmoidal curve for oxygen binding (Bohr, 1904). If the hemoglobin chains bound oxygen independently from one another then the binding curve would be hyperbolic like that for myoglobin. This method of oxygen transport would be less efficient, as a lower amount of oxygen could be released going from the environment of the lungs (95 mmHg) to the tissues (40mmHg). The middle part of the oxygen equilibrium curve is modelled by the Hill equation (Hill, 1910), for the reaction

$$Hb + nO_2 \leftrightarrow Hb(O_2)_n$$

$$\log\left(\frac{Y}{1-Y}\right) = n\log(P_{O2}) - n\log P_{50} \qquad \textit{Eq. 2.1}$$

where Y is the fraction of hemoglobin molecules saturated with $O_2$.

A Hill plot of log[(Y/1-Y)] versus log $P_{O2}$ approximates a straight line, with its slope at the binding midpoint (Y=0.5) called the Hill coefficient (*n*). The Hill coefficient is a measure of the cooperativity displayed in oxygen binding. Myoglobin displays non-cooperative binding and has a Hill coefficient of 1.0, whereas hemoglobin's binding is cooperative and has a Hill coefficient of 2.8.

## 2.5 ALLOSTERIC EFFECTS OF HEMOGLOBIN.

### *2.5.1 The Bohr effect.*

As the pH is lowered from 9.0 to 6.0 the oxygen affinity of hemoglobin is reduced, while its affinity for protons increases. This is called the alkaline Bohr effect (Bohr et al., 1904). Below pH 6.0 the oxygen affinity increases as the pH is lowered further. This observation is termed the acid Bohr effect.

This pH-dependent oxygen affinity allows hemoglobin to function in a dual role, as a respiratory carrier and to facilitate $CO_2$ transport by the uptake of protons. As $CO_2$ is expelled from the lungs there is an increase in pH and therefore a consequent increase in oxygen affinity, which favours the binding of oxygen.

$$CO_2 + H_2O \leftrightarrow H_2CO_3 \leftrightarrow H^+ + HCO_3^-$$

As the blood circulates through the capillaries $CO_2$ enters the blood, causing a drop in the pH which in turn lowers the oxygen affinity of the hemoglobin, resulting in the unloading of $O_2$ to the tissue.

A small amount of $CO_2$ (10%) becomes covalently bound to hemoglobin as a carbamino complex at the N-terminus (Ferguson and Roughton, 1934). The remainder of the $CO_2$ is transported as bicarbonate ions dissolved in the blood, with hemoglobin carrying the released protons. The bicarbonate is converted back to $CO_2$ in the lungs due to the release of protons from hemoglobin on oxygenation.

Various studies have been conducted on the native adult hemoglobin and on naturally occurring variants to determine which residues are important in the Bohr

effect. These residues (and their approximate contributions) are: alkaline Bohr effect, α1 Val N-terminal α-amino group (25%), α122 His (10%), β82 Lys (20%), β146 His (40%) and acid Bohr effect β143 His (50%) (Perutz et al., 1969; Kilmartin and Rossi-Bernardi, 1969; Perutz et al., 1980; Shih and Perutz, 1987; Perrella et al., 1994). It has however, been recognised that these contributions vary if other allosteric effectors, namely organic phosphates or chloride ions, are present (Rollema et al., 1975).

### *2.5.2 Allosteric effect of 2,3-bisphosphoglycerate*

The red blood cells of humans and most other mammals have a high concentration of 2,3-bisphosphoglycerate (2,3BPG). 2,3BPG is a glycolytic intermediate, with a concentration of around 5mM in cells. This concentration is roughly equivalent to that of the hemoglobin tetramer, and four times that of ATP (Bunn et al., 1977).

2,3BPG acts as an allosteric effector on hemoglobin. This molecule binds to the deoxyhemoglobin tetramer in a 1:1 ratio and lowers the oxygen affinity of the protein (Benesch et al., 1968a; Benesch et al., 1968b; Benesch et al., 1969). Under the same conditions no 2,3BPG is bound by oxyhemoglobin. There was compelling evidence that 2,3BPG bound at the N-terminal residues of the β subunits: Hemoglobins $A_{IC}$ and $F_1$ which are blocked at the N-terminus of their "β-like" chains have a lower affinity for 2,3BPG (Bunn and Briehl, 1970). Pyridoxal phosphate acts as an affinity label when added to hemoglobin and bound at the β N-terminus (Benesch et al., 1972). Perutz modelled the binding of 2,3BPG to deoxyhemoglobin and ascertained that the phosphate groups could bind to Val β1 N-terminal α-amino group, His β2, Lys β82 and His β143 of the two β subunits, bridging between them. He concluded that in oxyhemoglobin, in contrast, the β N-termini were too far apart for 2,3BPG to bind effectively (Perutz, 1970). Arnone determined the structure of deoxyhemoglobin with organic phosphate present and confirmed that the modelled site was indeed correct (Arnone, 1972).

2,3BPG thus stabilises the deoxy T state of hemoglobin, and increases the energy barrier for the T to R state transition, thereby decreasing the oxygen affinity.

### 2.5.3 The chloride effect

Chloride ions act as allosteric effectors, reducing the oxygen affinity of mammalian hemoglobins. This effect is more than double the alkaline Bohr effect (Perutz et al., 1994). Chloride ions lower the oxygen affinity of hemoglobin by stabilising the deoxy T state. It has been proposed that chloride ions diffuse into the central cavity down the two-fold axis of the tetramer, where they neutralise the excess positive charges contained there. The cavity is wider in deoxyhemoglobin allowing more ions to diffuse in. The chloride ions appear to be mobile and exhibit no significant electron density in X-ray structures of hemoglobin crystallized with chloride present (Kavanaugh et al., 1992; Perutz et al., 1993). The following positively charged residues are proposed to be important in the chloride effect: α99 Lys, α103 His, β1 Val N-terminal α-amino group, β2 His, β82 Lys, β104 Arg, β143 His, and β146 His. These cationic charges inside the cavity are partially countered by α94 Asp, α126 Asp and β101 Glu. The relative importance of each of these residues has been investigated by measuring the chloride effect in naturally occurring variant hemoglobins in which a substitution has occurred that reduces the cationic excess of the central cavity (Perutz et al., 1994).

The neutralisation of these excess positive charges in the larger internal central cavity of deoxyhemoglobin reduces the repulsive forces between the subunits and stabilises the T state hemoglobin. Furthermore it has been observed that chloride ions also alter the cooperativity of some variant hemoglobins (Bonaventura et al., 1994).

### 2.5.4 Nitric oxide binding

It has long been recognised that hemoglobin is able to bind nitric oxide (Keilen and Hartree, 1937). The heme iron has an affinity for $NO^{\bullet}$ that is at least 10 000 times greater than that for oxygen (Perutz, 1996). $NO^{\bullet}$, though an unstable and highly toxic radical, has been identified as an endothelial relaxing factor (Furchgott and Zawadzki, 1980; Palmer et al., 1987). Hemoglobin can bind NO both on the heme itself and at Cys β93, a residue conserved amongst all mammalian and bird hemoglobins. The binding of NO to hemoglobin is in equilibrium with NO bound to glutathione and other small thiols present in the red blood cell. It has been proposed that the mode of NO binding is dependent on the quaternary state of hemoglobin and this gives rise to the following mechanism of action. Deoxy T state hemoglobin

with NO bound on the heme enters the lungs where the presence of $O_2$ causes the allosteric transition to the R state and the transfer of NO to Cys β93. Blood flowing away from the lungs is exposed to a reduced oxygen pressure which promotes the T state of hemoglobin and effects NO release, some of which is recaptured by the hemes, and some of which is bound by small molecular weight thiols promoting dilation of the blood vessels, and enhancing the delivery of $O_2$ to the tissue (Gow and Stamler, 1998; Stamler et al., 1997; Jia et al., 1996).

## 2.6 HUMAN HEMOGLOBINS DURING DEVELOPMENT

As the very early human organism develops, the environment in which it grows changes, and so do its methods and problems of oxygen supply. Between weeks two and twelve of gestation, before the placenta is formed, three embryonic hemoglobins are produced: Gower I ($\zeta_2\varepsilon_2$), Gower II ($\alpha_2\varepsilon_2$) and Portland ($\zeta_2\gamma_2$) hemoglobins (Figure 2.3). Once the placenta has been established the embryonic hemoglobins are replaced by fetal hemoglobin. Shortly after birth fetal hemoglobin is replaced by the adult protein. The switching of hemoglobins produced coincides with the changes in the environment of the developing fetus. The early embryo scavenges oxygen from the mother's interstitial fluid. After approximately three months the embryo obtains oxygen from the mother's bloodstream via the placenta. Once born the child is able to utilise its own lungs for supplying oxygen.

With each phase of development; i.e. preplacental, placental and postnatal there is a different hemoglobin that appears to be adapted to function in the particular environment of the organism at that stage.

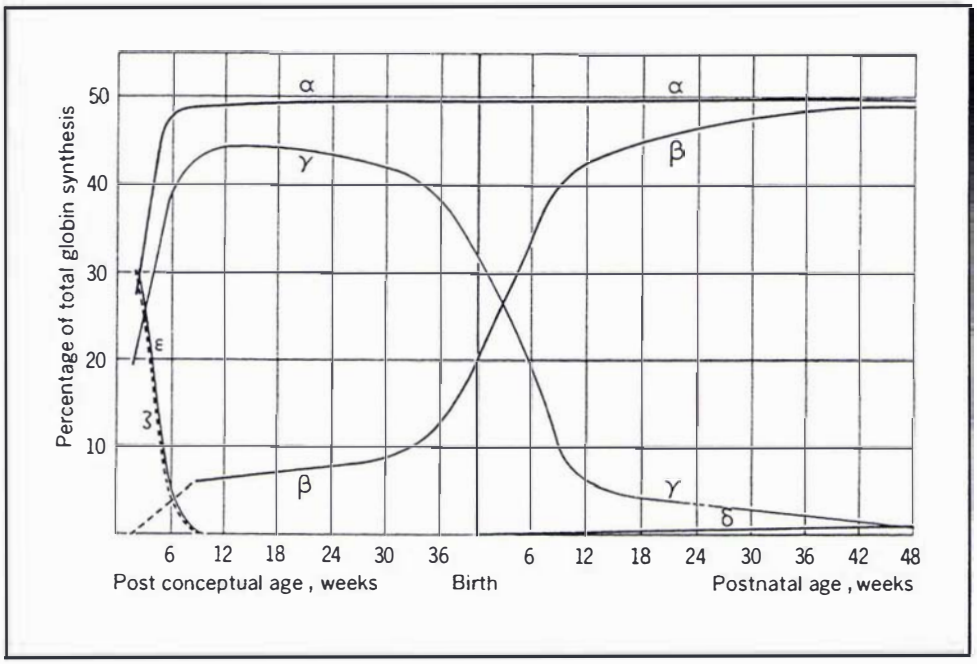



**Figure 2.3** Representation of the changes in human hemoglobin synthesis during prenatal and postnatal development.

Figure reproduced from Dickerson and Geiss, 1983.

### *2.6.1 Fetal hemoglobin*

Fetal hemoglobin (Hb F) has the subunit composition $\alpha_2\gamma_2$, thus having $\alpha$ chains as in adult hemoglobin, together with $\gamma$ chains. The $\gamma$ subunit comprises 146 residues like the $\beta$ chain, but differs in 39 positions (73% identity) (Schroeder et al., 1963). Two fetal hemoglobins have been characterised, i.e. $\alpha_2\gamma_2^A$ and $\alpha_2\gamma_2^G$ which differ by a single amino acid substitution, Ala → Gly at position 136 (Little et al., 1979).

In vivo, fetal hemoglobin has a greater oxygen affinity than that of the adult protein. This facilitates oxygen transfer across the placenta as oxygen is stripped from the adult hemoglobin present in the mother's blood to supply the fetus. However the intrinsic oxygen affinity of fetal hemoglobin, with no phosphate present, is actually lower than that of the adult molecule (Tyuma and Shimizu, 1970).

A comparison of the sequences of fetal and adult hemoglobin shows that 20 of the amino acid substitutions are found to occur on the surface of the molecule, whilst

10 are in the interior. There is an important amino acid replacement at position 143, β His → γ Ser. Residue 143 is one of the amino acids involved in binding the allosteric effector 2,3-bisphosphoglycerate. This substitution results in 2,3BPG being bound less tightly to fetal hemoglobin, thereby not lowering the oxygen affinity as greatly as in the adult protein.

The tertiary structure of fetal deoxyhemoglobin has been determined at 2.5 Å (Frier and Perutz, 1977). The β3 Leu → γ3 Phe substitution also inhibits the binding of 2,3BPG by shifting the N-terminus, thus reducing the effectiveness of γ2 His as a phosphate binding ligand. There are no significant changes in the α subunits. The sequence changes in the γ chain result in shifts in the N-terminus of the molecule with the A and E helices being brought closer together. The structural change is thought to be responsible for the lower intrinsic (i.e. with no phosphate present) oxygen affinity of fetal hemoglobin, as this conformation is similar to that of adult deoxyhemoglobin with 2,3BPG bound.

### *2.6.2 The embryonic hemoglobins*

There are three embryonic hemoglobins synthesised during the preplacental phase of human development; i.e. Gower I, Gower II, and Portland. The embryonic hemoglobins utilise the adult α chain and a similar ζ chain, together with the fetal γ chain and a novel ε chain, both of which are "β-like" subunits:

Gower I, $\zeta_2\varepsilon_2$

Gower II, $\alpha_2\varepsilon_2$

Portland, $\zeta_2\gamma_2$

In contrast to the adult hemoglobin, and to a lesser degree fetal hemoglobin, little is known about the functional and structural properties of the embryonic hemoglobins. This is mainly due to both ethical and practical problems of protein supply. Research has been undertaken on some animal models, e.g. chicken (Brown and Ingram, 1974), mouse (Bauer et al., 1975; Brittain et al., 1986) and crocodile (Grigg et al., 1993), but the relevance of this work to the human system is debatable (Mould et al., 1994).

Early research on the embryonic hemoglobins consisted predominantly of electrophoretic and denaturation studies. The appearance of an embryonic hemoglobin was first observed in small human embryos which had a hemoglobin with an alkali denaturation rate between that of adult and fetal hemoglobins (Drescher and

Kunzer, 1954; Kunzer, 1957). This observation appeared to be confirmed when a hemoglobin with a different electrophoretic mobility from the adult and fetal hemoglobins was discovered in the blood from a 20 week old fetus (Halbrecht and Klibanski, 1956). Electrophoretic studies on blood from a 10 week old embryo found two embryonic hemoglobins present, these being designated Gower I and Gower II (Huehns et al., 1961). The subunit composition of Gower II hemoglobin was determined by tryptic peptide mapping which showed that the adult α chain was present and a previously unknown subunit termed ε, giving Gower II the subunit composition $\alpha_2\varepsilon_2$. (Huehns et al., 1964). Gower I was incorrectly assigned the $\varepsilon_4$ quaternary structure and was thought to be the predominant embryonic hemoglobin (Hecht et al., 1966).

Another embryonic hemoglobin was identified in blood taken from an abnormal baby, in which synthesis of this embryonic hemoglobin had persisted until after birth. Electrophoresis showed that this hemoglobin had the subunit composition $x_2\gamma_2$ and it was named Portland 1 hemoglobin (Capp et al., 1967). Preliminary tryptic peptide patterns indicated that the x chain, now renamed ζ, was more closely related to the α chain than the β owing to a Tyr-Arg peptide that is unique to the α C-terminus, rather than the Tyr-His peptide which is present in the 'β-like' chains (Capp et al., 1970).

The primary structure of the ζ chain was determined by further trypsin and pepsin digests, and peptide fingerprinting. The resulting peptides were aligned with the α sequence. Only for positions 41-46 was no peptide match found (Kamuzora and Lehmann, 1975). A second sequence determination using peptide digestion led to completion of the analysis. This work revealed the total sequence, with a number of differences to the previously published results (Clegg and Gagnon, 1981). This protein sequence was confirmed as correct when the nucleotide sequence of the human ζ globin gene was determined (Proudfoot et al., 1982).

The sequence of the ζ chain is that of an 'α-like' globin chain. It contains 141 amino acids, with the characteristic Lys-Tyr-Arg C-terminus, and lacks the short D helix found in 'β-like' chains. There are 57 differences between the α and ζ sequences (60% identity), with several of these at positions of likely structural and/or functional importance (Figure 2.4). The N-terminal α-amino group of the ζ chain Ser is acetylated, in contrast to α1 Val. The substitutions at positions 20 (αHis → ζGln), 23 (Glu → Thr), and 50 (His → Pro) result in the probable loss of two intra-subunit

hydrogen bonds, and it is likely that the introduction of a Lys at position 82 will give an additional hydrogen bond to the heme propionate side-chains. There are three substitutions at residues involved in intersubunit interactions, i.e. 36 (Pro → His), 38 (Thr → Gln) and 119 (Pro → Ala), which may be responsible for altering the oxygen affinity of this hemoglobin (Clegg and Gagnon, 1981).

The ζ chain exhibits a greater sequence similarity to the two chicken 'α-like' globins than to the adult α chain (Melderis et al., 1974; Chapman et al., 1980) suggesting that the ζ subunit is some sort of "primitive" chain.

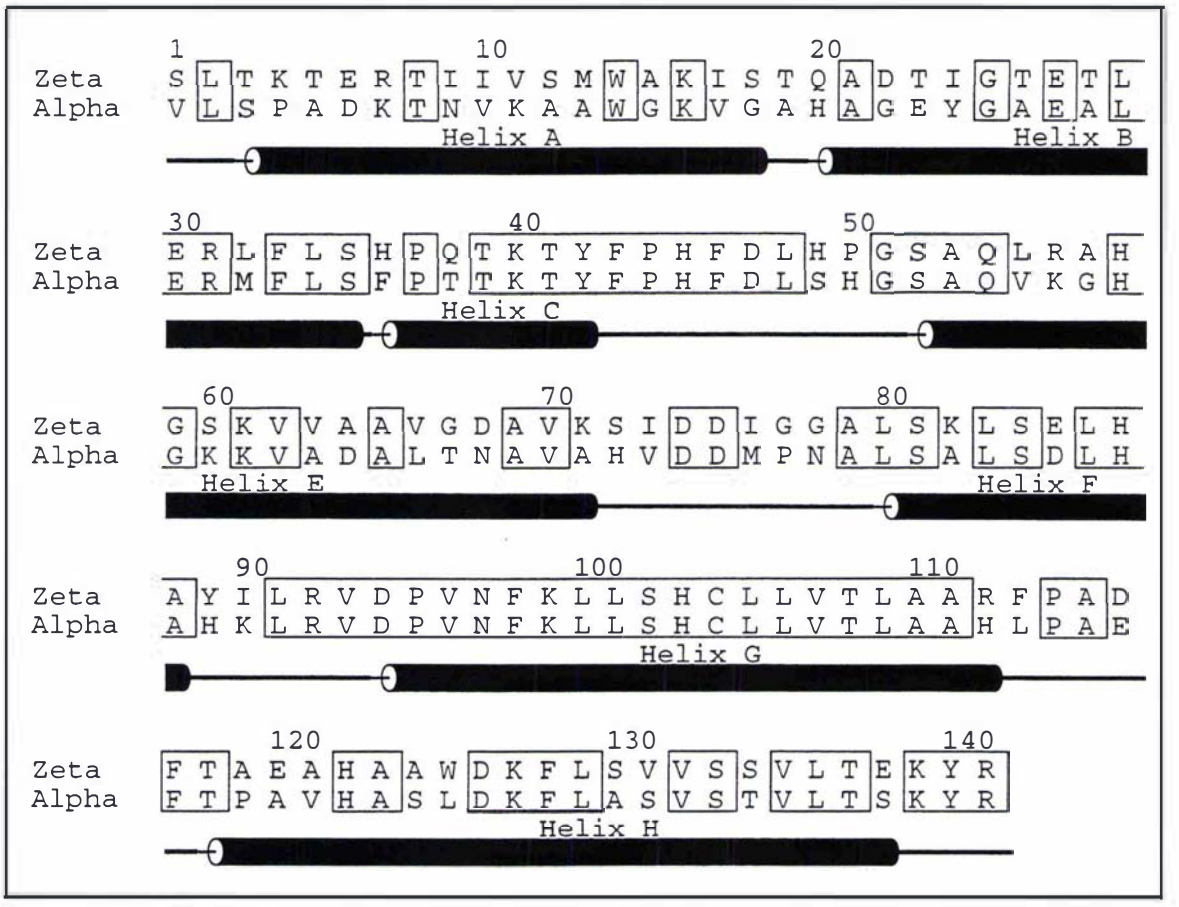



**Figure 2.4** Amino acid sequence alignment of the human hemoglobin ζ and α chains

Identical residues are boxed. Alignment displayed with ALSCRIPT (Barton, 1993).

Tryptic digestion and mapping analysis of the resulting peptides led to a proposed sequence for the ε chain, though the second half of the sequence was tentative (Gale et al., 1979). It was also established that the quaternary structure of Gower I

hemoglobin was in fact $\zeta_2\varepsilon_2$ and not $\varepsilon_4$ as first thought. The determination of the nucleotide sequence of the human ε globin corrected some discrepancies in the peptide sequence (Baralle et al., 1980). The ε chain is 146 amino acids in length, differing in 30 positions from the β chain (79% identity) and in 36 positions from the γ chain (75% identity). The majority of the key residues are conserved between the three 'β-like' chains (Figure 2.6). The ε chain retains the 2,3BPG-binding His143, in contrast to the γ subunit. As in the γ chain the amino acid at position 3 in the ε chain is a phenylalanine, whereas in the adult β chain this residue is a leucine. The increased side-chain size is thought to destabilise 2,3BPG binding by moving the His2 imidazole ring further away from the binding site, as proposed for Hb F (Frier and Perutz, 1977).

The evolutionary relationships and estimated time of divergence of the five human globin genes (α,ζ,β,ε,γ) have been calculated by examining the percentage divergence of amino acid replacements (Efstratiadis et al., 1980; Proudfoot et al., 1982). The relationships are summarised in Figure 2.5.

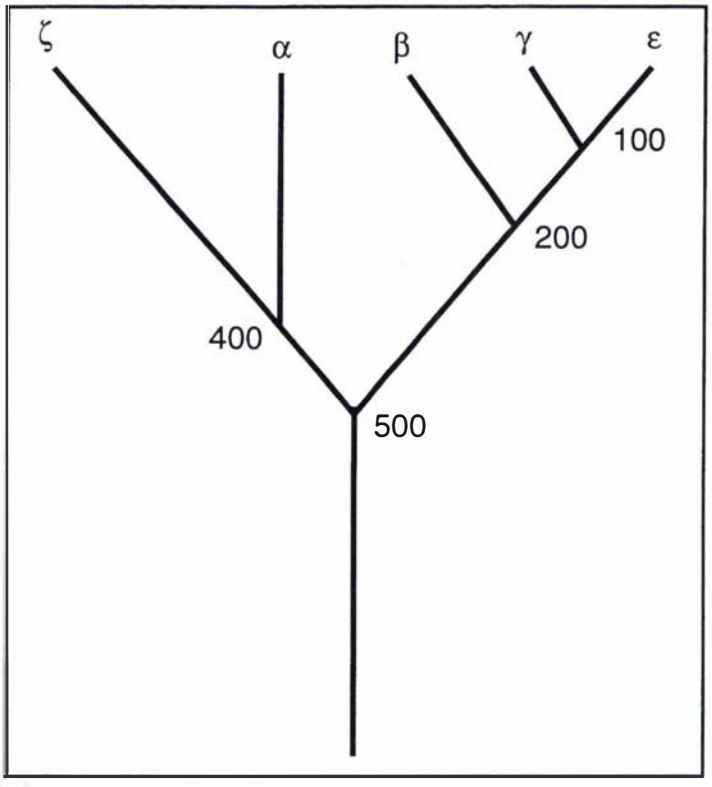



**Figure 2.5** An evolutionary tree of the human globin family

Numbers given are in units of millions of years. Figure modified from Proudfoot et al., 1982.

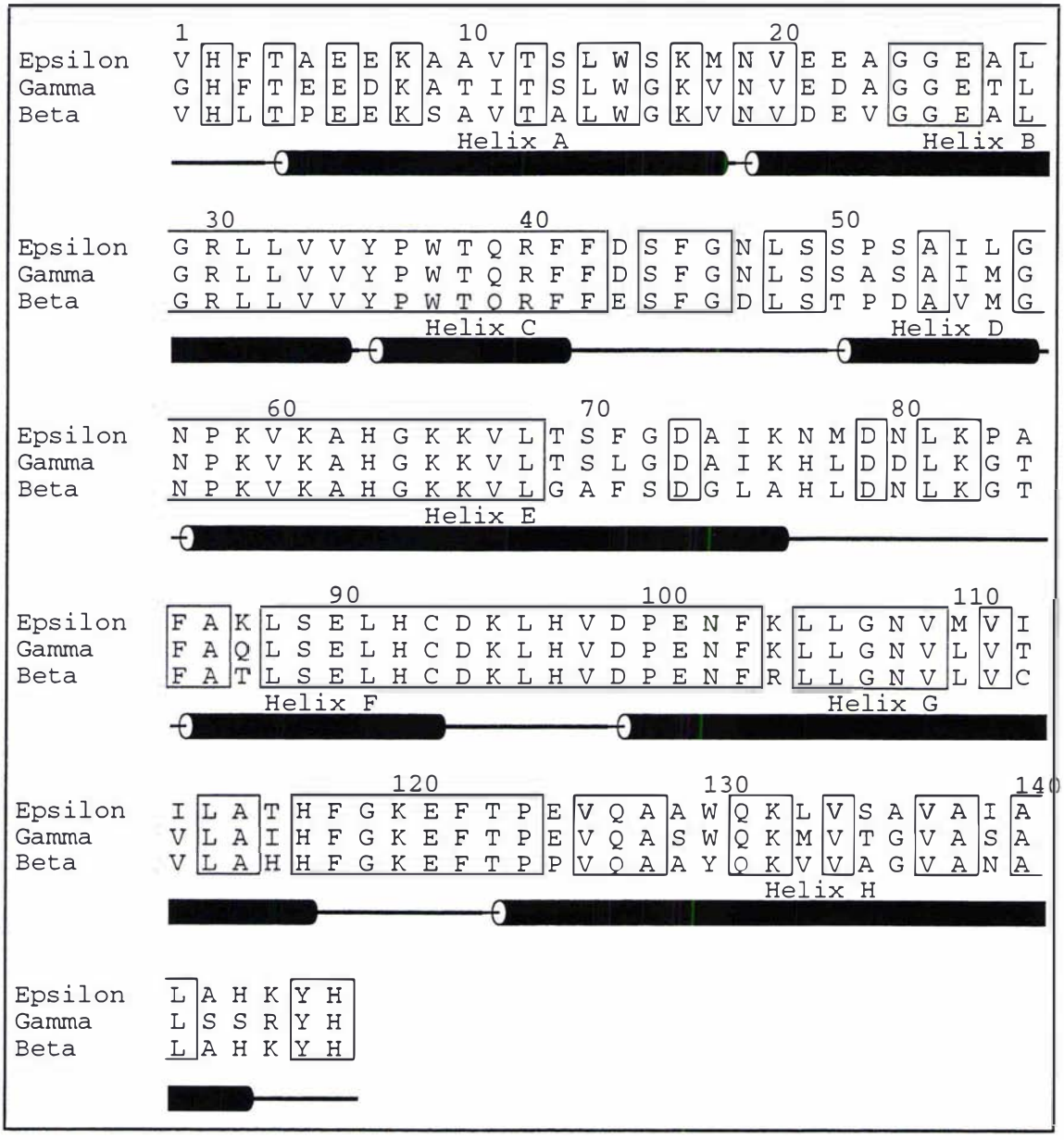



**Figure 2.6** Amino acid sequence alignment of the human hemoglobin 'β-like' chains ε, γ, and β

Identical residues are boxed. Alignment displayed with ALSCRIPT (Barton, 1993).

### *2.6.3 Genetic control of hemoglobin synthesis*

A large volume of research has been undertaken into investigating the switch from synthesis of fetal to adult hemoglobin in various animal models, with the switch considered an ideal system for study of how gene expression is regulated by devel-

opment (Stamatoyannopoulos et al., 1979). However, little research has been conducted on the timing of the embryonic globin switches and the change to synthesis of fetal hemoglobin.

Electrophoretic studies of hemoglobins present in early-stage embryos indicate that the embryonic ($\zeta$,$\varepsilon$) to fetal ($\alpha$,$\gamma$) globin switch starts around week 6 of pregnancy and is completed by week 12 (Gale et al., 1979).

This time-frame coincides with the site of erythropoiesis changing from the yolk sac to the liver, leading to the assumption that the embryonic hemoglobins are synthesised in the yolk sac, whereas the liver is responsible for the production of fetal hemoglobin. However the mechanism does not appear to be a straightforward developmental sequence of hemoglobin synthesis, as both adult and fetal hemoglobins have been found in embryos as young as six weeks old (Hecht et al., 1966). Further investigation was done by evaluating globin production in the erythroblasts from the yolk sac and liver of early embryos (Peschle et al., 1984; Peschle et al., 1985). This showed that the embryonic $\rightarrow$ fetal hemoglobin synthesis switch occurred as a gradual continuum. At 5 weeks only embryonic $\zeta$ and $\varepsilon$ globin genes are activated. By 6-7 weeks these genes are gradually suppressed, whilst the $\alpha$ and $\gamma$ genes are activated. By 7-8 weeks the $\varepsilon$ and $\zeta$ genes are completely inactive. These observations suggested that a single cell line is responsible, which is modulated through development to switch from embryonic then to fetal, and finally adult globin synthesis. Other research appears to support this idea (Wood et al., 1985). The alternative hypothesis is that a change in cell line occurs for each globin synthesis. The presence of some mixtures of globins in the blood at the same time could be due to cells undergoing the globin switch out of phase, as opposed to there being a cell line for each globin (Stamatoyannopoulos et al., 1987).

## 2.7 FUNCTIONAL PROPERTIES OF EMBRYONIC RED BLOOD CELLS

Until relatively recently, little was known about the functional properties of the embryonic hemoglobins, owing to both practical and ethical difficulties in obtaining sufficient quantities of the proteins to conduct these studies.

The oxygen dissociation properties of embryonic red cells were investigated by Huehns and Farooqui. They found that there was no apparent difference in the oxy-

gen affinity compared to fetal erythrocytes (Huehns and Farooqui, 1975). However as these measurements were carried out on whole cells, the relative globin concentrations present were unknown, hence correlations for the affinity of individual hemoglobin chains could not be assessed. The oxygen dissociation curve of Portland hemoglobin was measured, revealing a higher oxygen affinity compared to the adult protein. Portland hemoglobin also exhibited cooperative binding with a Hill coefficient of around 2, (compared to a coefficient of 2.8 for Hb A) and a diminished Bohr effect (Tuchinda et al., 1975).

It was not until recombinant embryonic hemoglobin could be expressed that other functional studies could be conducted. Wagenbach pioneered a method to produce adult hemoglobin from a single plasmid introduced into a yeast expression system (Wagenbach et al., 1991). This method was adapted to initially produce Gower II embryonic hemoglobin (Mould et al., 1994) (for a brief description see section 3.1 on page 34) and then the Gower I and Portland proteins.

## 2.8 OXYGEN BINDING AND ALLOSTERIC EFFECTS OF THE HUMAN EMBRYONIC HEMOGLOBINS

The production of recombinant embryonic hemoglobins has allowed for their physico-chemical and functional characterization.

### *2.8.1 Oxygen binding measurements*

Oxygen binding curves were measured for the three embryonic hemoglobins at various protein concentrations and under presumed physiological conditions (37°C, pH 7.4, 100mM NaCl) (Table 2.1, Hofmann et al., 1995b). All the embryonic hemoglobins have a significantly higher oxygen affinity than adult hemoglobin under these conditions and display cooperative binding with Hill coefficients ($n$) greater than 2.2.

**Table 2.1** Oxygen binding measurements for the embryonic and adult hemoglobins

| Hemoglobin | $P_{50}$* (mmHg) | n* |
|---|---|---|
| Adult $\alpha_2\beta_2$ | 9.8 | 2.9 |
| Gower II $\alpha_2\varepsilon_2$ | 5.5 | 2.4 |
| Gower I $\zeta_2\varepsilon_2$ | 3.5 | 2.3 |
| Portland $\zeta_2\gamma_2$ | 5.9 | 2.3 |

* Measurements made at pH 7.4 (50mM BIS-TRIS buffer), 37°C, 100mM NaCl

Analysis of oxygen binding curves measured in the absence of 2,3BPG show that all the embryonic hemoglobins follow the two-state model of allosteric action (Brittain et al., 1997). These calculations indicate that the T state of all three embryonic hemoglobins is considerably less stable than for the adult molecule. Hence it appears that the higher affinity and lower cooperativity of these hemoglobins relative to the adult protein arise from destabilization of the T state.

### *2.8.2 Allosteric effects*

#### *2.8.2.1 Binding of 2,3-bisphosphoglycerate*

The reduced binding of 2,3BPG to fetal hemoglobin is responsible for the higher oxygen affinity relative to the adult protein under physiological conditions (Tyuma and Shimizu, 1970). The oxygen affinities of the three embryonic hemoglobins were measured as a function of 2,3BPG concentration. As discussed previously (section 2.5.2 on page 16) 2,3BPG binds at a site formed between the β chains of the T state tetramer. All the residues that form this site in the adult β chain: Val 1 (N-terminal α-amino group), His 2, Lys 82 and His 143, are conserved in the ε chain suggesting that the Gower I and Gower II hemoglobins should bind 2,3BPG in a similar manner. It is noted that a substitution at position 3, β Leu → ε Phe, which also occurs in the γ chain of fetal hemoglobin, is present in the ε chain. The deoxy T state structure of fetal hemoglobin showed that the replacement of this leucine residue by the sterically more bulky phenylalanine resulted in a shift of the N-terminal A helix, causing the movement of β2 His away from the binding site. The binding of this effector is also reduced by the loss of two of its ligands (per tetramer) by the substitution: β143 His→ γ143 Ser. The γ chain is also found in the Portland embryonic hemoglobin and hence a drastically reduced affinity for 2,3BPG would be expected.

Measurements of the oxygen affinity of the three embryonic hemoglobins as a function of 2,3BPG concentration confirmed these hypotheses (Table 2.2). The ε chain containing Gower I and II hemoglobins exhibited a slightly reduced response to 2,3BPG, and the γ–containing Portland Hb showed a dramatically reduced response to the allosteric effector (Hofmann et al., 1995b).

**Table 2.2** 2,3-Bisphosphoglycerate binding to the embryonic and adult hemoglobins

| Hemoglobin | 2,3BPG Binding* |
|---|---|
| Adult $\alpha_2\beta_2$ | 0.45 |
| Gower II $\alpha_2\varepsilon_2$ | 0.5 |
| Gower I $\zeta_2\varepsilon_2$ | 0.5 |
| Portland $\zeta_2\gamma_2$ | 6.0 |

* Measurements made at pH 7.4 (50mM BIS-TRIS buffer), 37°C, 100mM NaCl. Apparent binding constant (mM) obtained from half-saturation point.

*2.8.2.2 The Bohr effect*

The oxygen affinity of the human embryonic hemoglobins with respect to changes in pH was investigated to measure the Bohr effect, both in the presence and absence of chloride ions. Gower II exhibited a similar response to the adult hemoglobin (Table 2.3). All the residues rationalised as being important in the adult hemoglobin Bohr effect (see section 2.5.1 on page 15) are conserved in this embryonic hemoglobin. The ζ chain of both Gower I and Portland embryonic hemoglobins contains the α1 Val → ζ1 Ser substitution, where the Ser has its α-amino group acetylated. This blocking of the N-terminal amino group would be expected to decrease the alkaline Bohr effect by approximately 25%. This could account for the reduced Bohr effect coefficients observed for these proteins experimentally.

**Table 2.3** Alkaline Bohr effect of human embryonic and adult hemoglobins

| Hemoglobin | Bohr coefficient in the presence of chloride# | Bohr coefficient in absence of chloride* |
|---|---|---|
| Adult $\alpha_2\beta_2$ | -0.45 | -0.23 |
| Gower II $\alpha_2\varepsilon_2$ | -0.45 | -0.25 |
| Gower I $\zeta_2\varepsilon_2$ | -0.17 | -0.10 |
| Portland $\zeta_2\gamma_2$ | -0.26 | -0.26 |

#Measurements made in 50mM BIS-TRIS buffer, 37°C, 100mM NaCl
*Measurements made in 50mM HEPES buffer
The Bohr coefficient was calculated from $\Delta\log P_{50}/\Delta pH$

*2.8.2.3 The chloride effect in the embryonic hemoglobins*

As described previously (section 2.5.3 on page 17) chloride ions act as allosteric effectors of hemoglobin without specific binding. The oxygen binding curves of the embryonic hemoglobins have been measured at varying concentration of chloride ions to determine their response to this effector. The degree of cooperative binding as expressed by the Hill coefficient remained unchanged. In the total absence of chloride, the embryonic and adult hemoglobins exhibited very similar oxygen affinities ($P_{50}$ values of 3.5 ±0.4mmHg). However, in the presence of varying concentration of chloride ions the relative change in oxygen affinity each of the four hemoglobins was quite different (Table 2.4).

**Table 2.4** Chloride sensitivity of the adult and embryonic hemoglobins

| Hemoglobin | Chloride Sensitivity* |
|---|---|
| Adult $\alpha_2\beta_2$ | 0.5 |
| Gower II $\alpha_2\varepsilon_2$ | 0.31 |
| Gower I $\zeta_2\varepsilon_2$ | 0.04 |
| Portland $\zeta_2\gamma_2$ | 0.0 |

* Measurements made at pH 7.4 (50mM HEPES buffer), 37°C.
The chloride sensitivity is expressed as $\Delta\log P_{50}/\Delta\log[Cl^-]$

The adult hemoglobin exhibited the greatest sensitivity to chloride ions and hence

displayed the largest decrease in oxygen affinity in the presence of this allosteric effector. The Gower II hemoglobin's response, though reduced in magnitude, exhibited a similar trend with chloride concentration. The Gower I hemoglobin displayed a markedly lower response, whereas the Portland hemoglobin's oxygen affinity was not modified by the presence of $Cl^-$ ions at all (Hofmann et al., 1995a).

These experimental observations have been rationalised in terms of the model proposed by Perutz (Perutz et al., 1993; Perutz et al., 1994) in which sequence substitutions that remove positive charges from the central cavity reduce the chloride effect. The large decrease in chloride sensitivity for Gower I hemoglobin and total removal for Portland hemoglobin suggests that at least two positive charges (per dimer) are removed from the central cavity. The following amino acid replacements in the ζ chain are thought to be important in the reduced chloride effect of the Gower I and Portland embryonic hemoglobins: α1 Val → ζ1 Ser-NAc, α138 Ser → ζ138 Glu, and for the Portland γ chain, β143 His → γ143 Ser, β125 Pro → γ125 Glu (Hofmann et al., 1995a).

The reduction in chloride effect, relative to the adult hemoglobin, for the Gower II protein is consistent with the removal of a single positive charge (per dimer) from the central cavity. This observation is not easily rationalised in terms of amino acid substitutions. Obviously the α chain is totally conserved between the two molecules suggesting that a substitution in the ε chain is responsible. Three possible substitutions between the β and ε chains have been suggested, 77 βHis → εAsn, 116 βHis → εThr and 126 βPro → εGlu (Hofmann et al., 1995a).

In order to investigate the effects of the N-terminal acetylation, a mutant Portland hemoglobin has been produced in which the acetylated ζ1 Ser is replaced with ζ1 Val, which does not become acetylated (Scheepens et al., 1995). The mutant protein exhibits a slightly lower oxygen affinity and an increased chloride effect, suggesting that the N-terminal α-amino group is partially responsible for chloride sensitivity.

### *2.8.3 Measurement of heme stability*

The relative binding constants for heme have been determined for each of the embryonic hemoglobins by the measurement of heme exchange from the appropriate methemoglobin to human serum albumin (Table 2.5), a reaction that can be fol-

lowed spectrophotometrically (Robson and Brittain, 1996). This exchange reaction only monitors transfer from the "β-like" chain, as the α heme is bound more than an order of magnitude more tightly (Winterhalter et al., 1971).

**Table 2.5** Equilibrium heme exchange for the "β-like" chains

| "β like" chain | Equilibrium Distribution Ratio* |
|---|---|
| β | 2.6 |
| ε | 0.1 |
| γ | 0.2 |

Measurements made at pH 9.0 (100mM TRIS buffer), 20°C.

*The equilibrium distribution ratios were calculated as [metheme albumin][apo Hb]/ [met Hb][albumin].

These measurements indicated that the γ chain binds heme ten times more tightly than the β subunit, and the ε chain binds heme 25 times more tightly than β.

### *2.8.4 Kinetic studies*

The association and dissociation kinetics of carbon monoxide binding have been investigated by stopped flow and flash photolysis experiments (Hofmann and Brittain, 1996). The measured oxygen association rates were found to be similar for the three embryonic and adult hemoglobins studied. Where the embryonic hemoglobins exhibited different properties to the adult molecule was in their kinetics of ligand dissociation. It appears that this could be responsible for the high oxygen binding affinities of the embryonic hemoglobins relative to the adult protein. The unusual Bohr effects of the embryonic proteins can also be rationalised by the pH dependence of the ligand dissociation rate.

The equilibrium constants for the tetramer-dimer reactions of the deoxy and oxygenated forms have also been investigated by these methods. The dimerization rate for the deoxyhemoglobin tetramers at pH 9.0 was found to be similar for each of the embryonic and adult hemoglobins.

## 2.9 AIMS OF THIS STUDY

For discussion of the human hemoglobin family it is convenient to separate human

development into 3 stages: 1) the preplacental stage that occurs very early in development, prior to the formation of the placenta, 2) the placental stage during which the placenta, a specialised organ that facilitates the transfer of essential molecules to the developing organism, is formed and 3) the post-natal stage during which the new-born child utilises its own organs for essential life functions. It appears that during each stage of development the embryo is subjected to particular problems of oxygen supply and that there are hemoglobin molecules synthesized that are optimised to transport oxygen under these different conditions. The three embryonic hemoglobins are expressed during the preplacental phase, fetal hemoglobin is then produced once the placenta is formed, and after birth adult hemoglobin is responsible for oxygen transport within the body.

An extensive body of research has been conducted on determining the functional, structural and mechanistic properties of adult hemoglobin, and a smaller but still substantial volume of work has been compiled on the fetal molecule. In contrast, little information has been obtained on the structure and function of the human embryonic hemoglobins until relatively recently.

The proposed function of the embryonic hemoglobins is that they are optimised for sequestering oxygen from the mother's interstitial fluid, implying that the embryonic protein would have a higher affinity for oxygen than the maternal adult hemoglobin. Studies of the functional properties of the embryonic hemoglobins have revealed differences in their oxygen binding behaviour in the presence of known allosteric effectors (section 2.8 on page 26). The differences in sensitivity to these allosteric effectors could fully explain the *in vivo* functional operation of these molecules (Hofmann and Brittain, 1996).

There has been considerable interest in the structure of hemoglobin since the pioneering work of Perutz, but except for human fetal deoxyhemoglobin (Frier and Perutz, 1977) no other structures of an embryonic or prenatal hemoglobin from any species has been determined. Structural analyses of the embryonic hemoglobins are needed in order to establish the structural basis for their altered properties. The present work was therefore undertaken as part of a programme aimed at analysing the structures of the whole embryonic hemoglobin family. The Gower II hemoglobin structure presented here is thus the first embryonic hemoglobin to be structurally characterised.

The structure determination of the Gower II human embryonic hemoglobin, and subsequent comparisons with the other human hemoglobins would provide insights into the structural origin of the observed functional differences.

The structure of the embryonic ε chain, and its interactions with the well characterised α chain would contribute to the general understanding of how the conformations of the amino acids contained in the "β-like" chains influence the functional properties of hemoglobin. Owing to the relatively large number of hemoglobin structures solved, and the high degree of structural and sequence conservation exhibited amongst the members of this protein family the Gower II hemoglobin structure would also add to the growing pool of information of how sequence determines three dimensional structure in this protein class. The evolutionary relationships of the members of the hemoglobin family have also been the subject of much study (Lesk and Chothia, 1980; Efstratiadis et al., 1980; Proudfoot et al., 1982; Aronson et al., 1994). Structural information, in conjunction with sequence data, is important for the study of evolution among a protein family and hence the structure of the ε chain would contribute to the understanding of the origin of these molecules.

Chapter 3

# Gower II Hemoglobin: Structure Determination

## 3.1 EXPRESSION AND PURIFICATION OF RECOMBINANT HUMAN EMBRYONIC HEMOGLOBINS

Recombinant Gower II hemoglobin was expressed in *Saccharomyces cerevisiae* and purified by Dr. T. Brittain, Dr. R. M. Mould and Dr. O. M. Hofmann (University of Auckland), as described earlier (Mould et al., 1994).

Briefly, cDNA for the embryonic ε globin gene was obtained from mRNA isolated from an erythroleukaemic cell line. The cDNA was amplified to high yield by PCR, and the globin gene ligated into Bluescript. This plasmid was used to transform *Escherichia coli* and subsequent transformed clones were selected and subjected to restriction analysis and automated sequencing to confirm the integrity of the ε globin gene. An adult (αβ) globin expression construct containing plasmid was modified by the excision of the α and β globin genes, followed by replacement with the ε gene and re-insertion of the α gene. Both these genes were placed under the control of galactose-regulated promoters. This resulting plasmid was then used to transform *Saccharomyces cerevisiae.* These cells were grown at 30°C on selective media and galactose introduced to induce expression of the αε genes. The cells were lysed after 50 hours incubation and the cell debris removed by centrifugation.

The embryonic Gower II hemoglobin was purified on CM-cellulose, Q-sepharose and Mono-S columns. The addition of exogenous heme proved unnecessary for hemoglobin production. The yield after 15 hours incubation was 20-30mg/L of cell culture. Higher yields could be obtained (up to 70 mg/L) but problems of sulphaem incorporation were encountered (Hofmann et al., 1994). The integrity of the Gower II hemoglobin was confirmed by UV/visible spectroscopy, SDS-PAGE analysis, gel exclusion chromatography and N-terminal protein sequencing.

## 3.2 CRYSTALLIZATION

### *3.2.1 Protein handling and storage*

All manipulation of the Gower II hemoglobin was undertaken in the presence of CO. The embryonic hemoglobins are all sensitive to the presence of oxygen, with binding of the ligand causing irreversible oxidation of the heme iron, producing methe-

moglobin, which has a tendency to precipitate. 2mM sodium dithionite was added to both the protein and precipitant solutions to aid in keeping the iron in the reduced state.

The protein was stored at -70°C, dissolved in 50mM BIS-TRIS buffer pH 7.4 and 100mM NaCl. Prior to crystallisation experiments the NaCl and buffer were removed by dialysing against milli-q water in a microcentricon. The protein concentration was determined by measuring the absorbance of the hemoglobin solution fully saturated with carbon monoxide ($\varepsilon_{1\%}$=8.4 at 540nm with a pathlength of 1cm).

### *3.2.2 Initial crystallization screening*

Several protein crystallisation screens that have been successful at identifying initial crystallization conditions with other proteins were tested on the Gower II hemoglobin (Kingston et al., 1994; Jancarik and Kim, 1991). Conditions from which the adult hemoglobin had been crystallised were also tried (Perutz, 1968). Vapour diffusion with hanging drops on plastic cover-slips suspended over Linbro tissue culture plates was the crystallisation method employed. All crystallizations were set up under an atmosphere of carbon monoxide in a glove box, and solutions were degassed and purged with carbon monoxide before use. Small bright red irregular clusters of crystals grew in 3µl drops by mixing equal volumes of the protein solution (40 mg/ml in 0.01M HEPES buffer pH 7.0 with 2mM dithionite) and a reservoir solution comprising 0.2M TAPS/KOH pH 8.5 and 21%(w/v) monomethylether-polyethylene glycol 5000 (MME-PEG 5000). Crystals appeared after 2-5 days equilibration against the reservoir solution at room temperature (Figure 3.1).

### *3.2.3 Crystallization by micro and macro-seeding*

Subsequent rounds of micro, followed by macro-seeding, were necessary to grow crystals large enough for data collection. Briefly, the initial crystals were crushed and used to streak-seed microcrystals into a 10 µl drop containing 15 mg/ml protein pre-equilibrated with 0.2M TAPS/KOH, pH 8.5, 14%(w/v) MME-PEG 5000. The drop was then left to equilibrate against a reservoir of 20%(w/v) MME-PEG 5000, 0.2M TAPS/KOH, pH 8.5. This gave small but nicely formed crystals (Figure 3.2); two or three subsequent rounds of macro-seeding with these crystals, using similar techniques, gave crystals large enough for data collection (Figure 3.3).

I am extremely grateful to Mrs. H. M. Baker for conducting these crystallisation experiments.

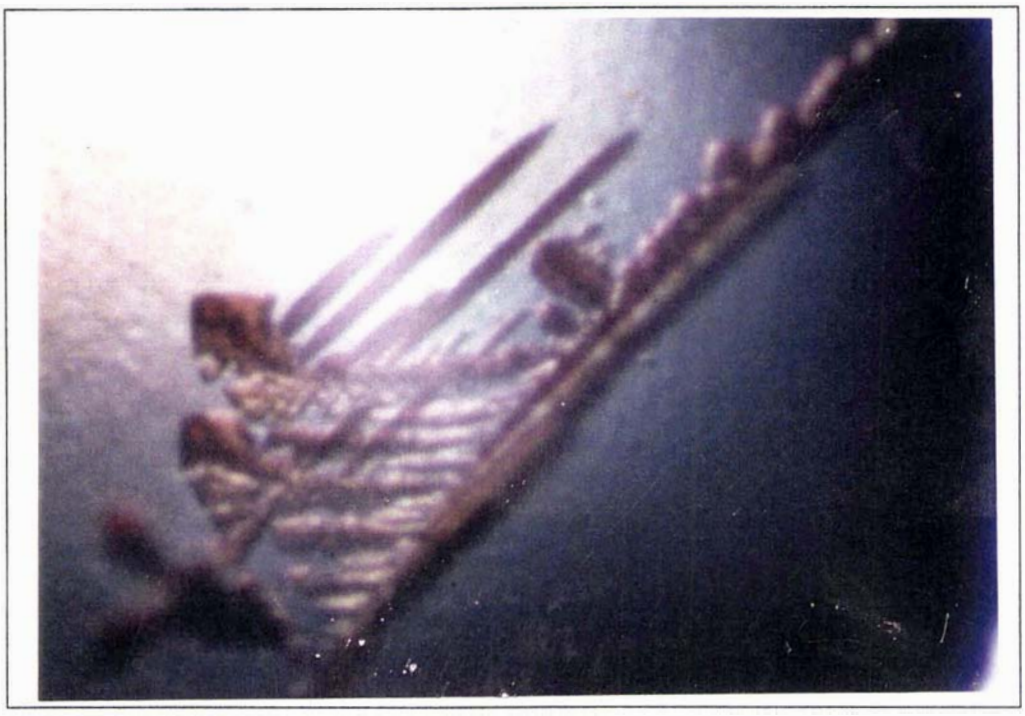


**Figure 3.1** Spontaneously nucleated Gower II embryonic hemoglobin crystals

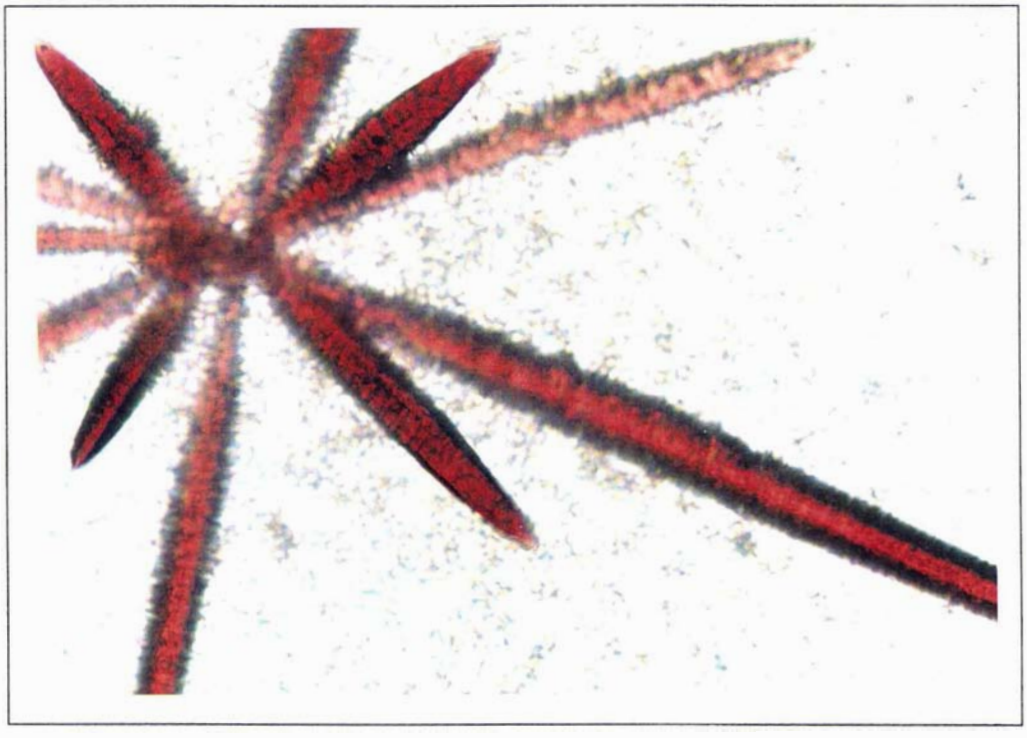


**Figure 3.2** Gower II hemoglobin crystals grown via microseeding

Magnification unrecorded for both these photographs

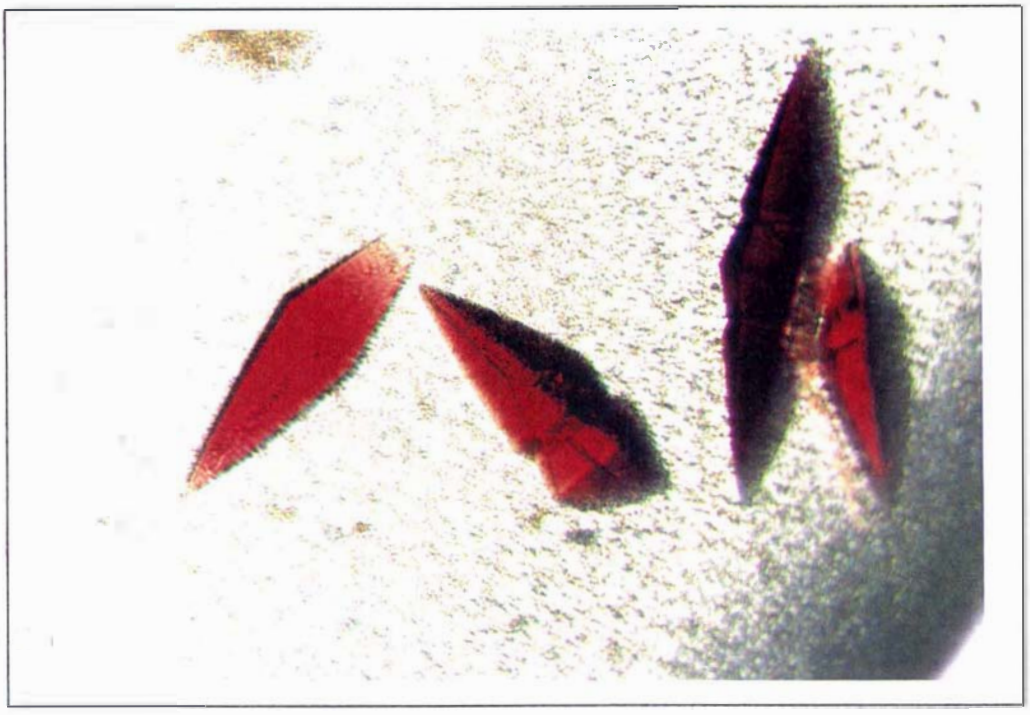


**Figure 3.3** Gower II hemoglobin crystals produced via macro-seeding

These crystals were typically of size 0.2 x 0.2 x 0.4 mm. Magnification unrecorded.

### *3.2.4 Further crystallization trials*

It transpired that these crystals (see Section 3.3.1) had an extremely long c axis (320 Å) that presented problems for data collection on the in-house generator and detector. The deoxy form of trout hemoglobin has been crystallized in a cell ($a$=$b$=63.3 Å, $c$=312.7 Å, $\gamma$=120°) of similar dimensions to the Gower II hemoglobin, though the symmetry of the cell is hexagonal and contains an $\alpha\beta$ dimer in the asymmetric unit (Tame et al., 1996). Further crystallization experiments were conducted in attempts to crystallize the Gower II hemoglobin in a different crystal form. These trials, however, were unsuccessful.

An analysis was made of the crystal contacts in the lattices of the previously crystallized human hemoglobins, in order to determine whether a particular substitution in the $\varepsilon$ chain may be responsible for the failure of the Gower II hemoglobin to crystallize in one of these lattices. No sensible rationale could be established however. Analyses of this kind are problematic; the manner in which protein sequence and structure affect crystallization behaviour is poorly understood, even for homologous

proteins. There are numerous examples (Day, 1993; N. Peterson, unpublished results) in which a single point mutation engineered into a protein, or the substitution of a metal ion at the active site (R. A. Edwards, unpublished results) prevents the protein crystallising in the same lattice as the wild type, even if the alteration occurs away from regions involved in crystal contacts. The most striking example of this is hemoglobin itself where the point mutation β6 Glu → Val in sickle cell hemoglobin (Ingram, 1956) completely alters the aggregation behaviour of the molecules so much that it affects the morphology of the red blood cell itself.

## 3.3 DATA COLLECTION

Two diffraction data sets were collected on the Gower II hemoglobin crystals with an R-axis IIC system, employing a Fuji imaging plate as an X-ray detector and CuKα radiation from a Rigaku RU-200 rotating anode generator. This data collection was carried out at Massey University.

### *3.3.1 Data set 1 (HBAE2)*

A crystal fragment was mounted in a wax-sealed glass capillary. A small amount of crystallisation mother liquor was included at the ends of the capillary to form a vapour front and prevent crystal dehydration, and the capillary sealed with wax. The crystal was oriented with the longest (needle) axis approximately pointing down the capillary, which was also the axis of rotation for the data collection. The generator was operated at 50kV and 100mA and the crystal to detector distance was set at 180mm for data collection at room temperature. Three still X-ray photographs were taken at phi angles of 0, 45, and 90° with an exposure time of 30 minutes. Diffraction spots were observed to about 2.9 Å, which at this crystal to film distance was close to the edge of the detector screen. The separation between the spots was small, but they could be resolved. Sixty oscillation photographs were taken with an exposure time of 50 minutes. The oscillation range for each photograph (Δphi) was 2°. The outer limit of the resolution decreased slightly as the data collection proceeded, presumably due to radiation damage of the crystal.

Automatic auto-indexing of the diffraction images using the programs REFIX (Kabsch, 1988) and the R-AXIS software (Sato et al., 1992) failed to produce a conclusive set of cell dimensions owing to the closeness of the reflections. The *a* and *b* cell lengths consistently converged to values around 62.5 Å, but the long *c* cell length varied greatly depending on which reflections were used in the computation. Consequently it was not possible to integrate the data successfully.

### *3.3.2 Data set 2 (HBAE4)*

The crystal was mounted in a glass capillary with the long needle axis approximately parallel to the capillary and rotation axes. The detector was set further back to 220mm to increase the spot separation. Four still photographs were taken at phi angles of 0, 20, 60, and 80° with an exposure time of 40 minutes, at room temperature. Very weak spots were present to 3.3 Å on the edge of the screen. Forty oscillation photographs were recorded, each with an exposure time of 60 minutes. The oscillation range for each photograph (Δphi) was 1.5°.

Autoindexing of the still photographs using the R-AXIS software gave a cell of:

$$a=62.8\ \text{Å},\ b=62.8\ \text{Å},\ c=320.9\ \text{Å},\ \alpha=\beta=\gamma=90°$$

The same cell was also calculated from autoindexing an oscillation photograph with REFIX. The intensity data were integrated with DENZO (Otwinowski, 1993) and then scaled, averaged and converted to structure factor amplitudes using the CCP4 suite of programs (Collaborative Computational Project, 1994) (results not shown).

The cell calculated above was used to process the first data set (HBAE2). Measurement of the profile-fitted intensities was conducted using the program DENZO. The measured intensities were scaled and merged in ROTAVATA and AGROVATA from the CCP4 suite. HKLVIEW from the CCP4 suite was used to display pseudo-procession photographs of the reciprocal lattice for data processed in P1. Two-fold symmetry was present about the *a* and *b* axes with odd *h00* (*h00*≠*2n*) reflections absent, indicating the presence of a two-fold screw axis along *a* (Table 3.2). The *hk0* section revealed the four-fold symmetry of the lattice. The *00l* reflections suggested that the *c* axis had either $4_1$ or $4_3$ symmetry (Table 3.1). Though no *0k0* reflections were measured, with the presence of point group 422 the *a* and *b* axes are identical necessitating that *b* must also be a two-fold screw axis, hence the lattice space group was assigned to be one of the enantiomorphs $P4_12_12$ (number 92) or $P4_32_12$ (number 96).

**Table 3.1** Subset of the *00l* Intensities for data set HBAE2

| *h* | *k* | *l* | Intensity | Sigma | I/Sigma |
|---|---|---|---|---|---|
| 0 | 0 | 69 | 20.5 | 23.2 | 0.9 |
| 0 | 0 | 70 | 32.0 | 16.0 | 2.0 |
| 0 | 0 | 72 | 1043.9 | 76.4 | 13.7 |
| 0 | 0 | 73 | -17.2 | 16.2 | -1.1 |
| 0 | 0 | 74 | -22.2 | 22.1 | -1.0 |
| 0 | 0 | 75 | -2.7 | 22.4 | -0.1 |
| 0 | 0 | 78 | 14.4 | 18.8 | 0.8 |
| 0 | 0 | 79 | -41.5 | 20.9 | -2.0 |
| 0 | 0 | 80 | 14011.1 | 653.6 | 21.4 |
| 0 | 0 | 90 | 6.7 | 16.4 | 0.4 |
| 0 | 0 | 91 | -4.8 | 22.8 | -0.2 |
| 0 | 0 | 92 | 721.6 | 51.6 | 14.0 |
| 0 | 0 | 93 | 5.8 | 16.1 | 0.4 |
| 0 | 0 | 94 | 4.7 | 22.1 | 0.2 |
| 0 | 0 | 95 | 3.9 | 15.7 | 0.2 |
| 0 | 0 | 96 | 54.3 | 22.1 | 2.5 |
| 0 | 0 | 97 | 23.8 | 21.7 | 1.1 |
| 0 | 0 | 98 | -13.6 | 20.7 | -0.7 |
| 0 | 0 | 99 | -3.5 | 20.7 | -0.5 |
| 0 | 0 | 100 | 15.3 | 20.8 | 0.74 |
| 0 | 0 | 101 | -6.4 | 20.1 | -0.3 |
| 0 | 0 | 102 | -2.3 | 19.0 | -0.1 |

**Table 3.2** Subset of the *h00* intensities for data set HBAE2

| *h* | *k* | *l* | Intensity | Sigma | I/Sigma |
|---|---|---|---|---|---|
| 2 | 0 | 0 | 3703.8 | 152.4 | 24.3 |
| 3 | 0 | 0 | -3.9 | 8.1 | -0.5 |
| 4 | 0 | 0 | 9699.4 | 432.1 | 22.4 |
| 5 | 0 | 0 | 8.6 | 18.9 | 0.5 |
| 13 | 0 | 0 | -0.4 | 30.5 | 0.0 |
| 14 | 0 | 0 | 4573.0 | 244.6 | 18.7 |
| 15 | 0 | 0 | -21.3 | 42.2 | -0.5 |
| 16 | 0 | 0 | 193.5 | 44.8 | 4.3 |
| 17 | 0 | 0 | 45.3 | 48.3 | 0.9 |
| 18 | 0 | 0 | 5246.1 | 200.8 | 26.1 |
| 19 | 0 | 0 | -83.3 | 54.2 | -1.5 |

The data were processed in space group $P4_32_12$ (number 96), with a total of 311 measurements being rejected due to failing the rejection criteria $( Ij - \langle I \rangle ) / \sigma(I) > 4$ (a summary of the data processing statistics is given in Table 3.3). Structure factor amplitudes were obtained from the measured intensities by the procedure of French and Wilson (French and Wilson, 1978), and put on to an approximate absolute scale by calculation of a Wilson plot (Wilson, 1942) using TRUNCATE from the CCP4 suite. This method forces all negative intensities to be positive, and inflates the weakest data which are likely to be underestimated. The Wilson estimation for the overall temperature factor was 62 $Å^2$.

**Table 3.3** Data processing statistics for the Gower II embryonic Hb data set HBAE2

| | HBAE2 Data# (40-2.9 Å) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Upper resolution limit (Å) | 8.96 | 6.42 | 5.26 | 4.57 | 4.09 | 3.74 | 3.46 | 3.24 | 3.06 | 2.90 | All |
| No. of measured reflections | 3041 | 5491 | 7072 | 8460 | 9498 | 10630 | 11599 | 12519 | 10342 | 7315 | 85967 |
| No. of unique reflections | 598 | 955 | 1186 | 1380 | 1534 | 1676 | 1813 | 1931 | 1943 | 1629 | 14645 |
| $\langle I \rangle / \langle \sigma(I) \rangle$ | 15.0 | 11.7 | 9.8 | 9.2 | 8.5 | 7.2 | 5.4 | 3.6 | 2.7 | 1.9 | 7.1 |
| Complete. (%) | 97.6 | 100 | 99.8 | 100 | 100 | 100 | 100 | 100 | 95.4 | 76.3 | 95.9 |
| Multiplicity | 5.1 | 5.7 | 6.0 | 6.1 | 6.2 | 6.3 | 6.4 | 6.5 | 5.3 | 4.5 | 5.9 |
| R-merge* (%) | 3.8 | 5.0 | 6.9 | 7.1 | 8.0 | 9.7 | 13.4 | 20.2 | 26.9 | 37.0 | 9.1 |

#Rejected measurements are not included in the statistics

*R-merge $= \Sigma_{hkl} \Sigma_j | I_j(hkl) - \langle I(hkl) \rangle | / \Sigma_{hkl} \Sigma_j | I_j(hkl) |$

Complete. is an abbreviation for completeness

The Matthews' coefficient (Matthews, 1968) gave a Vm of 2.47 $Å^3$/Da, assuming the presence of a tetramer in the asymmetric unit. This value is equivalent to the protein occupying 49% of the cell volume, or a solvent content of 51%.

A plot of the logarithm of the mean amplitude for reflections of constant *h*, *k* and *l* versus 1/resolution yielded three similar curves, indicating that the diffraction was isotropic. In this procedure all reflections of constant *h* are binned and the mean amplitude ($\langle F_{obs} \rangle$) and mean 1/resolution calculated. This procedure is repeated for *k* and *l* to give the final plot (Figure 3.4).

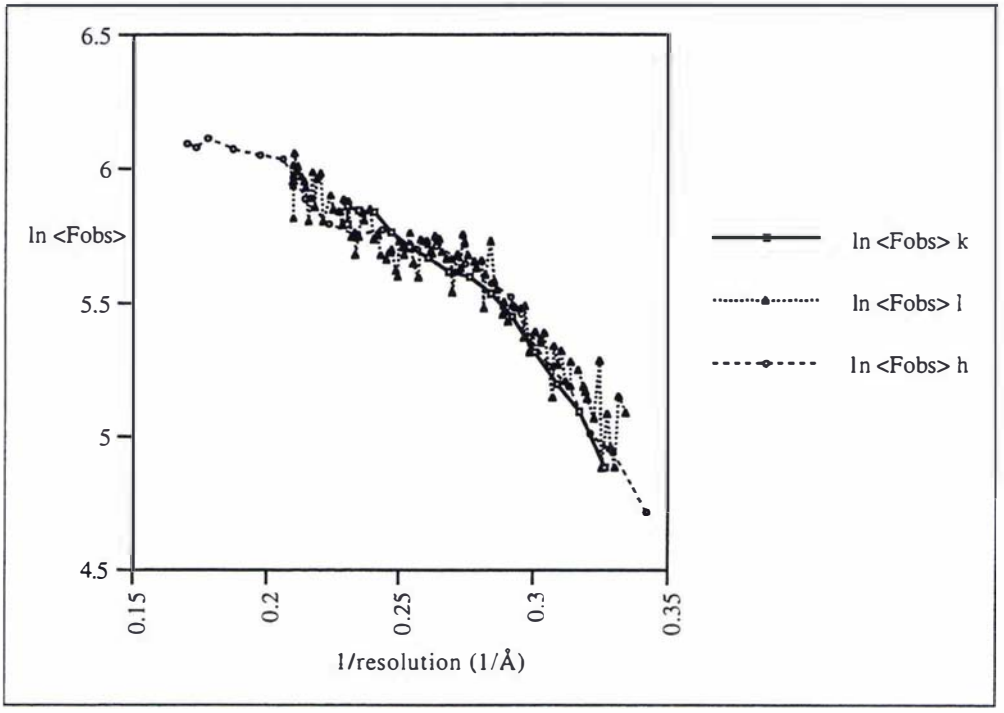



**Figure 3.4** Plot of the logarithm of the mean observed amplitude as a function of resolution for reflections of constant *h*, *k* and *l*

This figure was prepared with DATAMAN (Kleywegt and Jones, 1996).

In order to prevent overlap of the reflections during data collection owing to the long *c* axis of the crystal, it was necessary to set the detector a large distance back from the crystal resulting in decreased intensity of the measured observations. It was hoped that an improved data set could be collected on the more intense beam line of a synchrotron source. Two data sets were later collected at the BL-6A2 synchrotron facility at the Photon Factory, Tsukuba, Japan by Prof. E. N. Baker, Dr. H. R. Faber and Dr. R. L. Kingston. These crystals however did not diffract as well as those used at Massey, resulting in a data set to 3.4 Å.

For all molecular replacement calculations, refinement and map calculations the HBAE2 data set was used.

### *3.3.3 Self-rotation function*

A self-rotation function was calculated using GLRF (Tong and Rossmann, 1990)

which employs the reciprocal space method of Rossmann and Blow (Equation 3.1, Rossmann and Blow, 1962),

$$R(C) = \int_U P_1(X_1)\, P_2(X_2)\, dX_1 \qquad \textit{Eq. 3.1}$$

where the rotation function $R$ is calculated by comparing the Patterson function $P_1$ at position $X_1$ with the Patterson function $P_2$ at position $X_2$. In the case of the self-rotation function both $P_1$ and $P_2$ are calculated from the native structure factor amplitudes. The function is calculated over the spherical volume of Patterson space $U$. A maximum in the rotation function yields the relative orientation $C$ of identical molecules within the asymmetric unit.

The hemoglobin tetramer exhibits pseudo 222 point symmetry, having only one true two-fold owing to the non-equivalence of the α and ε chains. The self-rotation function was calculated using amplitudes of resolution 10.0 to 4.0 Å and magnitude greater than 6σ or absolute value 100. The outer Patterson integration radius was 20.0 Å and the Patterson origin peak was removed from the computation. The peak heights and a stereographic projection of the κ=180 section of the rotation function are given in Table 3.4 and Figure 3.5.

**Table 3.4** Self-rotation function calculation (GLRF) κ=180 section[#]

| Peak No. | Spherical Polar Angles (°) | | | Peak Height (σ) |
|---|---|---|---|---|
| | φ | ψ | κ | |
| 1 | 0.0 | 21.0 | 180 | 13.9 |
| 2 | 96.0 | 84.0 | 180 | 9.7 |
| 3 | 21.0 | 135.0 | 180 | 7.9 |
| 4 | 57.0 | 57.0 | 180 | 7.7 |

[#]Non origin peaks shown only

Polar angles (φ,ψ,κ) are defined such that φ is the angle from the cartesian x axis, and ψ as the angle from the y axis, with κ as the rotation about the axis defined by φ and ψ.

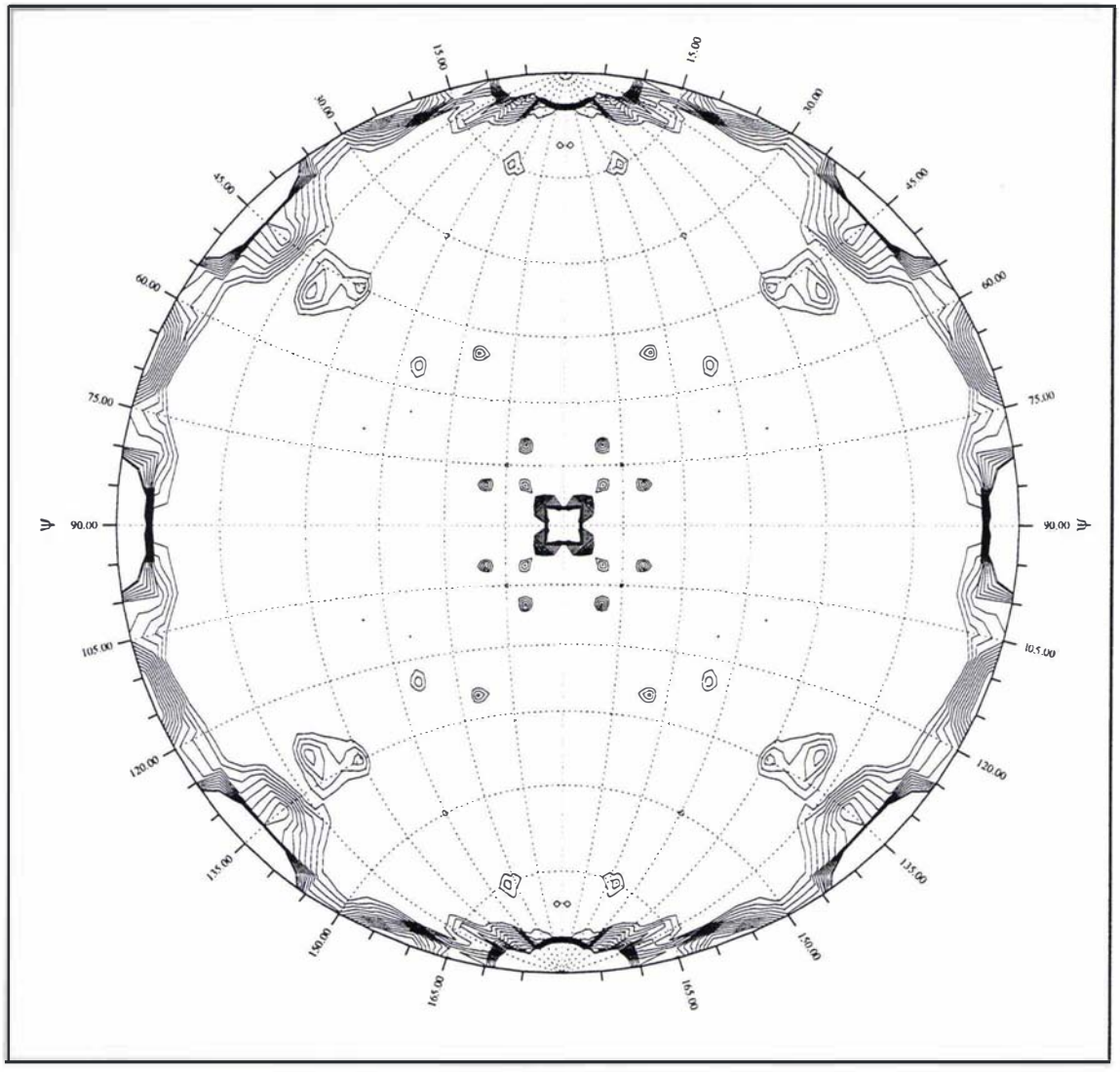



**Figure 3.5** Stereographic projection of the Gower II Hb data (HBAE2) self-rotation function κ=180 section

The plot of the self-rotation function has been contoured in steps of 0.8σ, starting at a value of 0.6σ above the mean value of the κ=180 section.

The highest peak on the κ=180 section (ϕ=0, ψ=25) suggested that the two-fold NCS operator between the two αε dimers lay in the AB plane, bisecting the crystallographic two-fold axes along the AB diagonal and either the A or B axis.

A native Patterson function was also calculated. The absence of any large non-origin peaks in the asymmetric unit indicated that there were no purely translational non-crystallographic symmetry operators in the cell.

## 3.4 MOLECULAR REPLACEMENT

### *3.4.1 Search model*

The model used for cross-rotation and translation searches was adult carbonmonoxy hemoglobin solved to 1.7 Å resolution (Silva et al., 1992, PDB code 1BBB). The whole tetramer, including the heme groups, was used as the search model. All solvent molecules were removed and the temperature factors were set to a value of 20 $Å^2$. The α subunits were left unaltered as this chain is totally conserved in the Gower II hemoglobin. In positions where the ε sequence differed from the β chain the side-chains were trimmed back to the common atoms, except for aspartate/asparagine, glutamate/glutamine, and valine/threonine substitutions which were considered to be similar enough to be left unaltered.

### *3.4.2 Rotation search*

Test calculations have shown that calculation of rotation functions using normalised structure factors, $|E|$, leads to peaks of improved signal to noise and a smaller dependence on the resolution limits of the data used, in comparison to structure factor amplitudes. Normalised structure factors are calculated (Equation 3.2) by dividing the structure factor amplitude, corrected for ε (a symmetry dependent correction) by its smoothed and interpolated average in equal volumes shells of reciprocal space (Driessen and Tickle, 1994).

$$|E|^2 = \frac{|F|^2/\varepsilon}{\langle |F|^2/\varepsilon \rangle} \qquad \textit{Eq. 3.2}$$

Normalised structure factors were calculated for the model in a large P1 cell (using the programs SFALL and ECALC from the CCP4 suite). The observed native structure factor amplitudes were also normalised.

A cross-rotation function was calculated utilising Crowther's fast rotation function algorithm (Crowther, 1972) as implemented in ALMN. Data between 10 and 4 Å and Patterson integration radii of 4 - 38 Å were used in the computation. Two unambiguous peaks related by a two-fold rotation were apparent at a height of 10σ, clearly distinct from the next highest peak at 4σ (Table 3.5).

**Table 3.5** Results of the rotation search (ALMN)

| Peak No. | Eulerian Angles (°) | | | Peak Height (σ) |
|---|---|---|---|---|
| | α | β | γ | |
| 1 | 18.2 | 42.7 | 209.5 | 10.4 |
| 2 | 71.5 | 42.6 | 151.1 | 9.9 |
| 3 | 69.7 | 53.0 | 346.7 | 4.0 |

Eulerian angles (α, β, γ) are defined such that α is a rotation of the coordinate system about **z**, β is a rotation about the new y axis, and γ is a rotation about the new z direction.

The same two peaks were consistently present as the top distinct peaks in the rotation function, with resolution limits changed to 8.0-4.0 Å and 12.0-4.0 Å, and Patterson integration radii changed to 4-40 Å.

The two solution peaks are related by a two-fold rotation, described in polar angles as ω=70°, ϕ=49°, and κ= 180°, where ω is the angle from the z axis, ϕ is the angle from y, and κ the rotation about this axis. This orientation was present as the third highest peak in the self-rotation function.

### *3.4.3 Translation search*

TFFC from the CCP4 suite was used to calculate a translation function. This program calculates a full-symmetry (Harada et al., 1981) formulation of the Crowther and Blow T2 function (Crowther and Blow, 1967). The translation function calculation would resolve the space group ambiguity. A similar position was found for both space groups, but the height of the peak in space group 96 indicated that this was the most probable symmetry (Table 3.6 and Table 3.7).

**Table 3.6** Translation function results (TFFC) -- space group $P4_32_12$ (96)

| Peak No. | Fractional Coordinates | | | Peak Height (σ) |
|---|---|---|---|---|
| 1 | 0.22209 | 0.39991 | 0.18591 | 37.4 |
| 2 | 0.22264 | 0.39971 | 0.13750 | 19.5 |

**Table 3.7** Translation function results (TFFC) -- space group $P4_12_12$ (92)

| Peak No. | Fractional Coordinates | | | Peak Height (σ) |
|---|---|---|---|---|
| 1 | 0.22177 | 0.40071 | 0.18584 | 19.2 |
| 2 | 0.22171 | 0.40047 | 0.43575 | 19.1 |

The next highest noise peaks in the translation function had the same x and y coordinates, but were smeared along the z axis.

The translation function calculations were repeated for all other space groups of point group 422. The only other space group with a solution significantly above the noise level was $P4_122$ (Table 3.8). In this space group x and y fractional coordinates for the position of the tetramer are identical to that found for $P4_32_12$, but the molecule is translated along z by 12.5% of the cell *c* axis. This shift results in steric clashes of the symmetry related molecules making the molecular replacement solution untenable. A space group of $P4_122$ is also not consistent with the systematically absent reflections observed, but it is intriguing that there exists a set of intermolecular cross vectors in this space group that are similar to those of the true lattice. This result is unexplained at this stage.

**Table 3.8** Translation function results (TFFC) -- space group $P4_122$ (91)

| Peak No. | Fractional Coordinates | | | Peak Height (σ) |
|---|---|---|---|---|
| 1 | 0.22217 | 0.39997 | 0.06095 | 21.1 |
| 2 | 0.30306 | 0.32032 | 0.06092 | 10.0 |

The rotation and translation searches were repeated with several calculations using AMORE. (Navaza, 1994). Structure factor amplitudes of several resolution ranges were used in the computations: 14.0 - 3.4 Å, 10.0 - 4.0 Å, and 8.0 - 4.0 Å and outer Patterson integration radii of 31 Å and 34 Å for the rotation function. Again two unequivocal peaks were observed, indicating that in this instance, with such a distinct molecular replacement solution, normalising of the structure factors proved to be unnecessary. The translation function was calculated in space groups

$P4_32_12$ and $P4_12_12$. In space group $P4_32_12$ the same orientation and position was determined as that from ALMN and TFFC. After the rigid-body fitting algorithm of AMORE this solution for the tetramer had a correlation coefficient of 56.1% and an R-factor of 41.1% (for data in the range 10.0 to 3.5 Å), with no other solutions present with a correlation coefficient greater than 10% and an R-factor less than 57%. When the translation function and rigid-body fitting were calculated in space group $P4_12_12$ the highest peak had a correlation coefficient of 15% and an R-factor of 55%. These results from the translation function calculations unambiguously defined the space group as $P4_32_12$.

In order to determine whether the quaternary structure of the tetramer was biased towards the search model the molecular replacement calculations were repeated using an αβ dimer (of the tetramer described previously) as the search model. In order to facilitate interpretation of the results of the calculation, the dimer was orientated and positioned by the molecular replacement solution matrix as determined for the tetramer. The calculations were conducted with AMORE in space group $P4_32_12$ utilising the same parameters as described previously, except for a reduction in the Patterson integration radius to 17 Å due to the smaller size of the search model. The molecular replacement calculations positioned the two dimers to form the same tetramer as located previously. The molecular replacement results after rigid-body fitting are given in Table 3.9.

**Table 3.9** αβ dimer molecular replacement solutions

| Rotation Function Solutions (Eulerian angles) (°) | | | Translation Function Solutions (Fractional Coordinates) | | | Correlation Coefficient* (%) | R-factor* (%) |
|---|---|---|---|---|---|---|---|
| α | β | γ | x | y | z | | |
| 357.9 | -0.11 | 1.26 | 0.9935 | 0.0058 | -0.002 | 47.5 | 46.8 |
| 57.27 | 38.84 | 223.33 | 0.5203 | 0.5629 | 0.4312 | | |

*for data in the range 10.0 to 3.5 Å

### *3.4.4 Validation of the molecular replacement solution*

The molecular replacement solution was unambiguous. The same solution was obtained using two different programs (ALMN/TFFC and AMORE) and with normalised and unnormalized structure factors. The top peak was consistent when the

resolution limits of the data used in the calculations were changed, and the Patterson radius for the rotation function computation was varied. The same solution was obtained using an αβ dimer as the search model. The molecules pack well in the crystal, forming a sensible crystal lattice (Figure 3.6) and as the refinement progressed, chemically reasonable crystal contacts were apparent. From calculation of SIGMAA-weighted Fourier maps (Read, 1986), phased from the molecular replacement solution, several of the omitted side-chains were clearly visible in the difference density (Figure 3.8, on page 55).

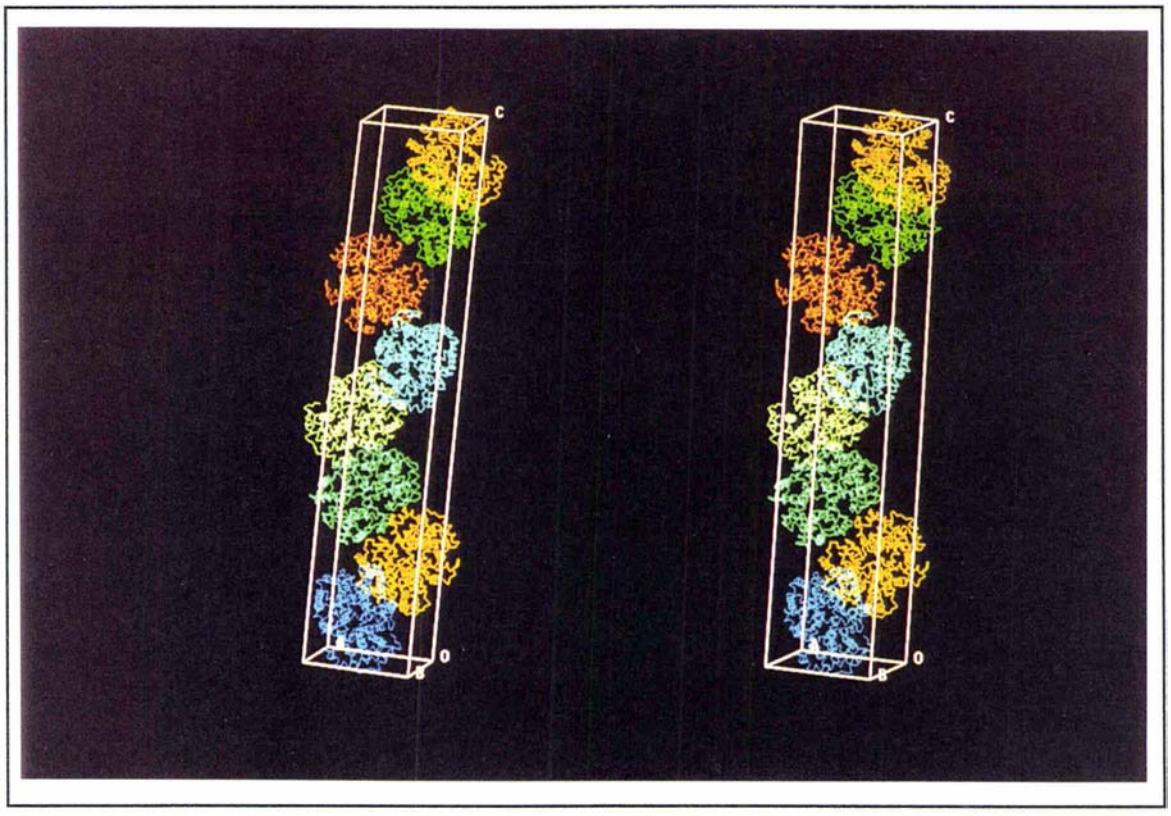



**Figure 3.6** Crystal packing of the Gower II embryonic Hb molecules

This figure were prepared with TURBO-FRODO (Roussel and Cambillau, 1991)

A retrospective analysis of the self-rotation function shows that the NCS two-fold is present on the κ=180 section as the third highest peak. Analysis of the crystal packing yielded no obvious explanation of why the large peak at ϕ=0, ψ=25 dominated the κ=180 section.

In order to determine whether this NCS operator was accounted for in the model a self-rotation function was computed using structure factors calculated from the

refined Gower II hemoglobin model ($F_c$). The parameters for the calculation were the same as for the observed native data self-rotation function. A listing of the peak heights and a stereographic projection of the $\kappa$=180 section are given in Table 3.10 and Figure 3.7.

**Table 3.10** Self-rotation function calculation (GLRF), calculated structure factors $\kappa$=180 section[#]

| Peak No. | Spherical Polar Angles (°) | | | Peak Height ($\sigma$) |
|---|---|---|---|---|
| | $\phi$ | $\psi$ | $\kappa$ | |
| 1 | 0.0 | 20.0 | 180 | 13.9 |
| 2 | 125.0 | 45.0 | 180 | 9.7 |
| 3 | 21.0 | 135.0 | 180 | 7.9 |
| 4 | 57.0 | 57.0 | 180 | 7.7 |

[#]Non origin peaks shown only

Polar angles ($\phi$,$\psi$,$\kappa$) are defined such that $\phi$ is the angle from the cartesian x axis, and $\psi$ as the angle from the y axis, with $\kappa$ as the rotation about the axis defined by $\phi$ and $\psi$.

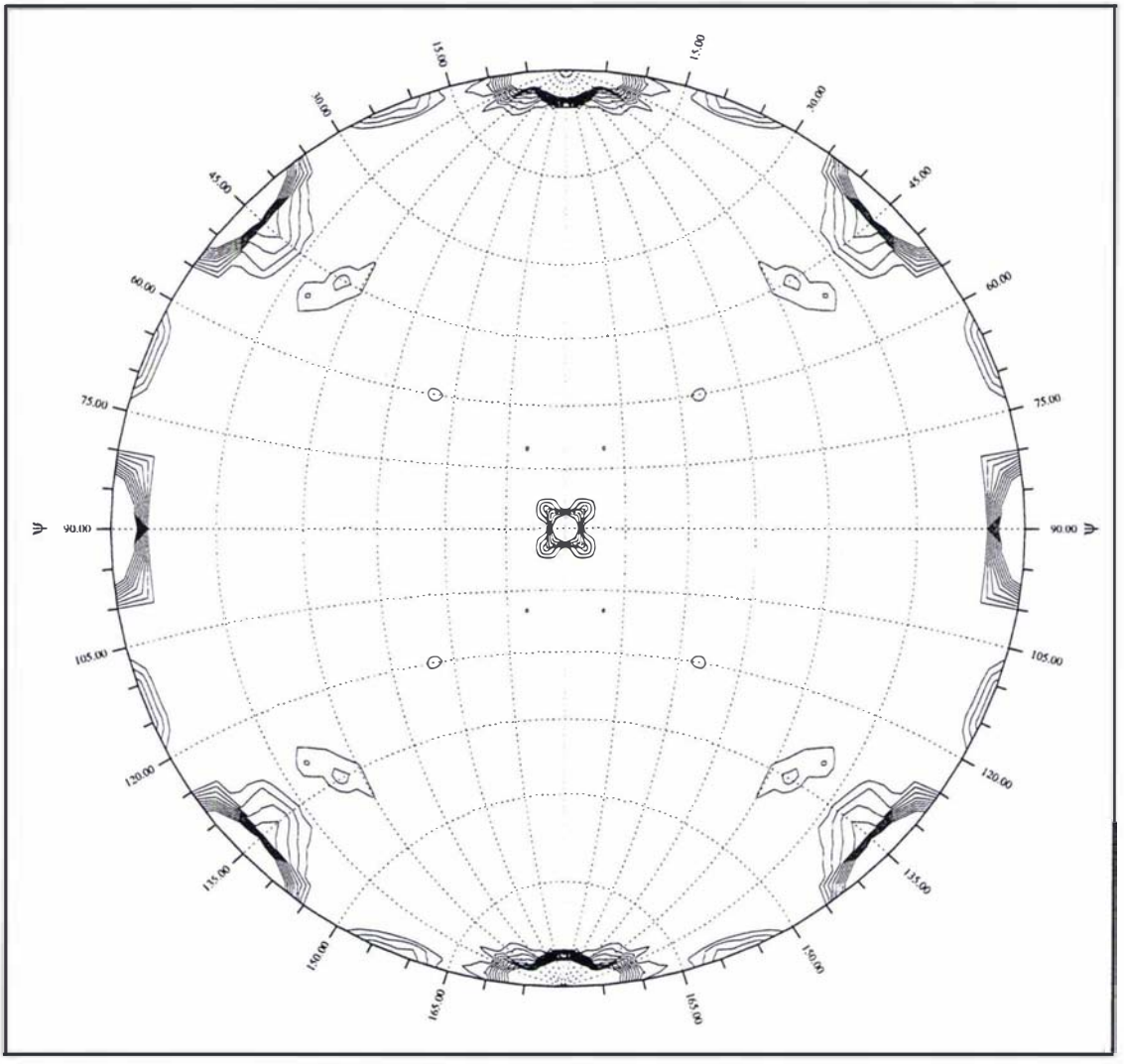



**Figure 3.7** Stereographic projection of $F_c$ self-rotation function κ=180 section.

The plot of the self-rotation function has been contoured in steps of 0.75σ, starting at a value of 17σ above the mean value of the κ=180 section.

### *3.4.5 Crystal contacts*

The crystal contacts are formed predominantly between the $\alpha_1$ subunit and the other three subunits of crystallographically-related tetramers (Table 3.11). It is interesting to note that the $\alpha_2$ heme propionate is involved in making a hydrogen bonded crystal contact.

**Table 3.11** Potential inter-tetramer hydrogen bonds

| Source Atom | Target Atom | Distance (Å) | Symmetry Operation | Translation (Tx Ty Tz) |
|---|---|---|---|---|
| $\alpha_1$12 Lys NZ | $\alpha_2$90 Lys O | 3.40 | 1/2+x, 1/2-y, 1/4-z | 0 0 0 |
| $\alpha_1$16 Lys NZ | $\alpha_2$ Heme O2D | 3.10 | 1/2+x,1/2-y, 1/4-z | 0 0 0 |
| $\alpha_1$44 Pro O | $\varepsilon_2$59 Lys NZ | 3.20 | -y, -x, 1/2-z | 1 1 0 |
| $\alpha_1$45 His N | $\varepsilon_2$44 Ser OG | 3.29 | -y, -x, 1/2-z | 1 1 0 |
| $\alpha_1$116 Asp OE2 | $\alpha_2$61 Lys NZ | 3.13 | 1/2+x, 1/2-y, 1/4-z | 0 0 0 |

## 3.5 REFINEMENT OF THE GOWER II HEMOGLOBIN MODEL

### *3.5.1 Overview of the Gower II hemoglobin refinement*

The adult hemoglobin search model provided a good molecular replacement solution for the calculation of an initial phase set. However the final stages of the refinement were not straight forward, principally due to the moderate resolution of the data. With the availability of only a modest number of reflections for refinement, and the absence of any independent phase information, care was necessary in order to prevent biasing the model when rebuilding. It was also important not to overfit the data by the introduction of unwarranted numbers of extra parameters. The magnitude of the free R-factor, as well as the difference between this statistic and the conventional R-factor were monitored in an effort to prevent this. In order to prevent the introduction of a large number of parameters by the modelling of individual solvent atoms, a bulk solvent model was used for the Gower II hemoglobin refinement. Owing to the twofold symmetry of the Gower II hemoglobin tetramer it was possible to reduce the number of refinable parameters by incorporating this NCS information in the form of constraints and restraints on the refinement. Considerable effort has been undertaken in recent years to improve not only the algorithms and protocols for protein structure refinements but also the methods of assessing the accuracy of protein structures (Dodson et al., 1996; Brunger, 1992a; Branden and Jones, 1990; Kleywegt and Brunger, 1996). Several of these improved methods have been incorporated in the Gower II hemoglobin refinement and are discussed below.

### *3.5.2 Cross-validation during refinement*

It is widely accepted that the preferred method for following the course of a macromolecular refinement is by the free R-factor (Brunger, 1992a). The free R-value is

calculated from a subset of reflections that have been excluded from the refinement. This yields a statistic that is independent from the refinement, but allows the process to be monitored. The omitted set should contain at least 1000 reflections for the free R-value to be a statistically valid value (Dodson et al., 1996). Approximately 10% (1485 reflections) of the HBAE2 data were randomly selected for the free-R set. These free reflections were conserved throughout the refinement, with the exception of the initial refinement in TNT, where no automatic facility for the calculation of a free R-value existed. Simulated annealing in X-PLOR was used to remove any bias from the model towards the free reflections after these refinement steps (Kleywegt and Brunger, 1996). A solution to this problem was found for later refinement in TNT, so that a free R-factor could be calculated.

The presence of NCS in the crystal results in correlations between reflections in reciprocal space. Consequently there may exist relationships between reflections in the working and test sets giving a free R-value that is no longer strictly independent from the refinement process, and that is likely to be underestimated. The refined structure of the MS2-RNA complex which exhibits ten-fold NCS has a free-R of 0.209 and R-factor of 0.192 (Valegard et al., 1994). The structure of glucose fructose oxidoreductase from *Zymomonas mobilis* has six copies in the asymmetric unit, some of which are related by nearly pure translational NCS symmetry (Kingston et al., 1996). The free R-factor and conventional R-factor for the refined structure were 0.22 and 0.21 respectively. In these extreme cases of high NCS it has been suggested that the free-R set of reflections be assigned in thin randomly selected resolution shells so that all correlated reflections end up in either the working set or the test set. It is not clear what effect omitting complete resolution shells of reflections will have on the refinement process and map calculations. Test cases with a molecule with four-fold NCS revealed that although the absolute value of the free-R (with randomly selected reflections) was lower than for a similar structure without NCS, the free R-factor was still a valid indicator for the refinement (Kleywegt and Jones, 1995).

### *3.5.3 Model building methods*

After rigid-body refinement, maps with SIGMAA-weighted Fourier coefficients, $(2m|F_o| - D|F_c|)$ and $(m|F_o| - D|F_c|)$, (where $m$ is the figure of merit, and $D$ is an estimate of the error in the structure from coordinate errors) were computed (Read, 1986), and electron density maps were calculated using phases calculated from the model. All reflections were included in the map calculations. At several

positions where the β and ε sequences differed, and hence the side-chain had been truncated to alanine, there were peaks in the difference density map indicating the position of the omitted side-chain (Figure 3.8).

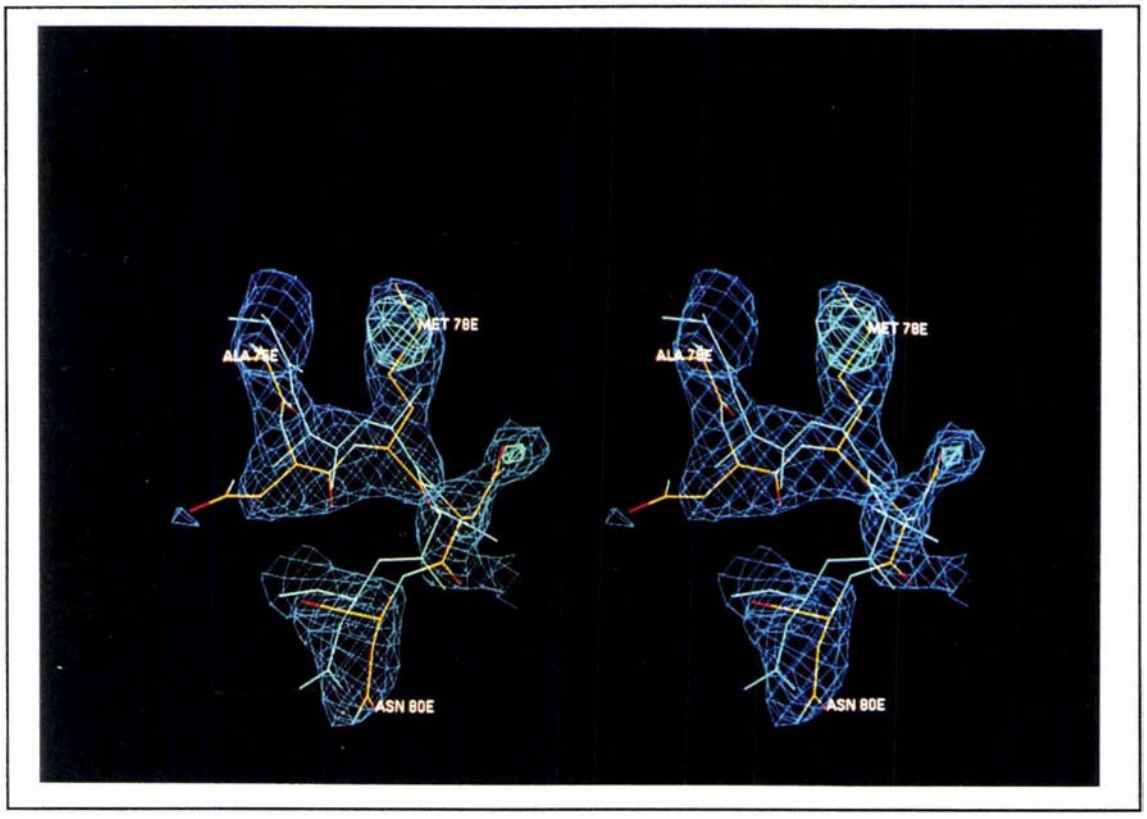



**Figure 3.8** Electron density maps of ε Met 78

The $2m|F_o| - D|F_c|$ is shown in dark blue and is contoured at 1.0σ, with the $m|F_o| - D|F_c|$ map, in light blue, contoured at 3.0σ. The phases are calculated from the MR solution model

The molecular replacement solution model (with residue 78 as Ala) is shown in light blue. The refined coordinates are shown coloured by atom type (C=yellow, N=blue, O=red, S=green).

The presence of these peaks further validated the molecular replacement solution. The side-chains that differed in the ε chain and were visible in the maps were added to the model, and other alterations to the model were made using the program TOM (Jones, 1982). After subsequent rounds of least squares or maximum likelihood refinement, maps as described above were calculated, and the model was rebuilt, initially using TOM and later using TURBO-FRODO (Roussel and Cambillau, 1991). The maps calculated were generally of high quality, though owing to the limited resolution of the data it was difficult to determine peptide orien-

tations and side-chain torsion angles in some regions of the structure. In these ambiguous areas the structures of the other human hemoglobins were used as guides in the model building, in addition to the considerable database of information on protein stereochemistry available in the model building programs and literature (e.g. Ramachandran plot and rotamer libraries).

### *3.5.4 Treatment of NCS during the refinement*

The major problem of protein structure refinement is the lack of measurable observations, when compared to the total number of parameters to be refined. For the Gower II hemoglobin refinement the current model of the tetramer in the asymmetric unit contains 4386 atoms. Positional refinement only of these atoms yields 13158 parameters, whilst the data set collected to 2.9 Å contains 15641 reflections. If an individual isotropic temperature factor is refined for each atom the number of parameters exceeds the number of reflections. The most powerful way of reducing the number of parameters refined is by the use of NCS. Molecules related by NCS can be constrained to be identical, resulting in a reduction of the number of parameters by a factor of the order of the NCS. This of course assumes that all the molecules in the asymmetric unit are identical, which at medium to low resolution appears to be a reasonable assumption where there is not enough data to model any differences. Analysis of protein structures containing NCS that had been refined at high resolution with no NCS restraints or constraints imposed, show that the average rms difference of the Cα atoms between the monomers is ~0.4 Å (Kleywegt, 1996). An additional advantage of employing NCS constraints is that the amount of manual model building is reduced, as only one monomer in the asymmetric unit has to be adjusted, though it is necessary to check the fit of the other monomers to the maps to ascertain that the NCS is indeed obeyed.

The alternative method of treating NCS is to restrain the molecules to be similar, with a weighting scheme determining how closely the monomers are related. The reduction of the number of parameters refined in this case is dependent on the weights chosen. This method is more flexible in allowing different parts of the molecules to be restrained with different weights.

A mixture of NCS constraints and restraints was used during the refinement of the Gower II hemoglobin structure. During the simulated annealing refinement the model was subjected to very tight NCS restraints, resulting in rms differences (for

all atoms) between the two αε dimers of 0.025A. During the TNT least squares and later maximum likelihood refinement NCS constraints were imposed. The NCS operator was calculated by least squares superposition of the αε dimers (Kabsch, 1976). At various stages during the refinement cycles of rigid-body refinement were conducted to refine the position of the monomers, and the NCS operator was subsequently recalculated. During the last round of refinement in X-PLOR the NCS constraints were relaxed and NCS restraints were imposed to enable side-chains that were ordered in only one dimer to be modelled.

### *3.5.5 Solvent model and bulk solvent correction*

Protein crystals typically contain 30-70% solvent (Matthews, 1968) and hence the ordered water molecules generally modelled for protein structures represent only a small fraction of the total solvent scattering in the crystal. Owing to the limited resolution of the data no water molecules were included in the atomic model of the Gower II hemoglobin, hence it was considered more important that the scattering contribution due to the bulk solvent be taken into account. If this component of the scattering is ignored the calculated structure amplitudes are systematically greater than the observed amplitudes at low resolution, creating problems in scaling these terms for refinement and map calculations. This effect is due to an overestimation of the electron density contrast at the protein surface if the macromolecule is placed in a 'vacuum', instead of disordered solvent.

One solution to this problem is to omit the low resolution terms below 5 Å. This protocol is unsatisfactory, however, since the omission of these reflections causes distortions in Fourier maps (Kostrewa, 1997).

Two methods for correcting for the bulk solvent contribution to the scattering were employed in the refinement of the Gower II hemoglobin. The bulk solvent correction implemented in TNT (Tronrud et al., 1987), REFMAC (Murshudov et al., 1997), and SHELXL (Sheldrick, 1997) was first employed by Moews and Kretsinger (Moews and Kretsinger, 1975). This protocol employs an application of Babinet's principle, in which the scattering due to the bulk solvent ($F_{sol}$) is assumed to have an amplitude directly proportional to that from the protein ($F_{prot}$), with opposite phase (Equation 3.3).

$$F_{sol} = -K_{sol}F_{prot}\exp\left(-B_{sol}\frac{\sin^2\theta}{\lambda^2}\right) \qquad \textit{Eq. 3.3}$$

and hence the total scattering ($F_{total}$) is the sum of the protein ($F_{prot}$) and bulk solvent ($F_{sol}$) contributions (Equation 3.4):

$$F_{total} = F_{Prot} - K_{sol}F_{prot}\exp\left(-B_{Sol}\frac{\sin^2\theta}{\lambda^2}\right) \qquad Eq.\ 3.4$$

$K_{sol}$ is a scale factor which is dependent on the relative electron density of the bulk solvent when compared to the protein (with values typically 0.7 - 0.9), and $B_{sol}$ is the temperature factor of the disordered solvent (with values typically 100 - 300 Å$^2$). These bulk solvent scale factors are dependent on the scaling parameters $K$ and $B$ which bring $F_{obs}$ on to an absolute scale. During the TNT refinement it was found that it was not possible to simultaneously refine all four of these scaling parameters with 2.9 Å resolution data. The bulk solvent scale factors were varied systematically and the free R-factor was monitored to determine which values yielded the best scaling model. The best parameters were $K_{sol}$=0.7 and $B_{sol}$=300 Å$^2$, though the free R-factor was not sensitive to small changes in these values.

An alternative approach is the bulk solvent correction as implemented in X-PLOR and CNS (Jiang and Brunger, 1994). A mask is placed around the protein by rolling a solvent probe over its Van der Waals surface. The region outside the mask is assigned to be bulk solvent and filled with a uniform value of electron density. Structure factors for the solvent contribution ($F_{sol}$) can be calculated via Fourier transform. These structure factors are modified by a scale factor $K_{sol}$ and temperature factor $B_{sol}$, and vectorially added to the structure factors calculated from the protein atoms, $F_{prot}$ (Equation 3.5).

$$F_{total} = F_{prot} + K_{sol}F_{sol}\exp\left(-B_{sol}\frac{\sin^2\theta}{\lambda^2}\right) \qquad Eq.\ 3.5$$

The values of $K_{sol}$ and $B_{sol}$ are determined by least squares minimization of the total model structure factors versus the observed data. In X-PLOR it was possible to refine the bulk solvent parameters successfully. This bulk solvent model makes no assumption about the relationships of the phases between the protein and bulk solvent structure factors.

Tests were conducted comparing the two solvent corrections described, as implemented in the programs TNT and X-PLOR. These comparisons are not strictly of just the solvent corrections, as this component of the scaling cannot be isolated from the scaling of $F_{obs}$. The same coordinates (final refined TNT coordinates) were used to calculate the statistics shown in Table 3.12.

**Table 3.12** Statistics for the implementation of the two bulk solvent models in the Gower II Hb refinement

| Bulk Solvent model | Free R-factor | R-factor |
|---|---|---|
| TNT: Babinet | 0.273 | 0.234 |
| X-PLOR: Babinet | 0.273 | 0.225 |
| X-PLOR: masked | 0.239 | 0.197 |

The values of $K$ and $B$ were refined as well as the solvent correction. In the two Babinet-type bulk solvent treatments the solvent parameters were set to values of $K_{sol}$=0.7 and $B_{sol}$=300 Å$^2$

Test calculations comparing the two bulk solvent models described have been conducted on *Eco*RV endonuclease, leading to the conclusion that the mask-derived bulk solvent correction as implemented in X-PLOR is the best available (Kostrewa, 1997). It has been shown that the assumption made in the Babinet-type solvent correction, i.e. that the phases derived from the bulk solvent density are opposite to those from the protein electron density is only true at resolutions lower than 15 Å (Urzhumtsev and Podjarny, 1996). This is presumably a consequence of the assumptions inherent in the application of Babinet's principle: that the two complementary masks (in this case the protein and bulk solvent regions) are comprised of the same scattering matter and that the electron density distributions for the two regions are smooth and featureless. The calculations conducted on the Gower II hemoglobin appear to support this conclusion, with the free R-value being substantially lower for the X-PLOR bulk solvent correction. However the maps were not dramatically improved, indicating that this drop in free R-value did not appear to correlate with an improved phase distribution. A comparison of two maps calculated from the same atomic model but from structure factors incorporating the two different bulk solvent corrections yielded a correlation coefficient of 0.98.

Similar results were found during the refinement of phosphoserine aminotrans-

ferase from *E. coli* (G. Hester, personal communication). The structure was refined at 2.8 Å resolution with NCS restraints. The crystallographic R-factors for the two bulks solvent models described above were: REFMAC Babinet-type solvent model, R-free = 0.322 and R = 0.265, and X-PLOR mask-type solvent correction R-free = 0.286 and R = 0.227. The maps calculated from each set of structure factors did not appear to differ greatly, though extensive tests were not conducted.

### *3.5.6 Temperature factor model*

There has been considerable debate in the literature as to what constitutes the most suitable temperature factor model for protein structure refinement at a particular resolution (Dodson et al., 1996), with the refinement of each macromolecule having its own peculiarities of data quality (independent of resolution) and the presence or absence of NCS. A suggested method for determining an appropriate B factor model is to start with a simple model that refines few parameters e.g. grouped temperature factors with one B per residue refined, and then introduce more sophisticated treatments whilst monitoring the free R-value to determine whether the introduction of more parameters is justified (Kleywegt and Brunger, 1996).

Refinement of temperature factors was begun after the simulated annealing refinement in X-PLOR with the free R-factor at 30.1%. Initially one temperature factor per residue was refined in TNT. However by monitoring the R-free, the best temperature factor model was achieved by refining individual isotropic temperature factors using the 'knowledge based' restraints suggested by Tronrud (Tronrud, 1996a). However a problem arises in determining what is a reasonable drop in the free R-factor, indicating an improvement in the model, when the difference between the free R-factor and the conventional R-factor increases at the same time, indicating overfitting in the refinement.

The statistics shown below (Table 3.13) are calculated for the last round of refinement in TNT using the maximum likelihood target. In each case the temperature factors were reset to 50 $Å^2$ and positional and temperature factor refinement was conducted with NCS constraints and a bulk solvent correction (as described previously). Though the figures given below were for the last round of refinement in TNT, these B factor models were also tested in this manner at various other stages of the refinement.

**Table 3.13** R-factors for TNT temperature factor models

| B factor model | Free R-factor | R-factor |
|---|---|---|
| Grouped B: One B per residue | 0.286 | 0.252 |
| Grouped B: Two Bs per residue | 0.280 | 0.243 |
| Restrained individual isotropic | 0.270 | 0.225 |

During the final refinement rounds conducted with X-PLOR the NCS constraints were removed for the last refinement round and replaced with NCS restraints. SInce this resulted in the introduction of double the number of parameters (though tightly restrained) it was deemed necessary to re-evaluate the temperature factor model by examination of the free R-factor as a function of the different temperature factor models. The results of these calculations are given in Table 3.14

**Table 3.14** R-factors for X-PLOR temperature factor models

| B factor model | Free R-factor | R-factor |
|---|---|---|
| Grouped B: Two Bs per residue | 0.235 | 0.186 |
| Restrained individual isotropic | 0.230 | 0.174 |

The refinement with restrained individual isotropic temperature factors has converged to a model with the lower R-free value. Again, however, the difference between the R-factor and the free R-factor has increased, suggesting that the model has been somewhat overfitted. The rms difference for the temperature factors between bonded atoms was 4.2 $Å^2$, a value usually considered higher than desirable at this resolution. For these two reasons the grouped temperature refinement was considered the better protocol, and the final model for Gower II hemoglobin has two temperature factors refined per residue, one for main-chain atoms and another for side-chain atoms. The heme has one temperature factor for the prosthetic group, and another for the bound ligand.

### *3.5.7 Progress of the Gower II hemoglobin refinement*

*3.5.7.1 X-PLOR rigid-body refinement.*

Rigid-body refinement of the initial model, from molecular replacement, was carried

out using X-PLOR (Brunger, 1992b) with each of the four subunits allowed to move independently. The R-factor decreased from 0.395 to 0.360 with a corresponding drop in the free R-factor from 0.388 to 0.350 during 40 cycles of refinement with data between 15.0 and 4.0 Å. The rotations of the subunits were less than 2°, with translations less than 0.7 Å. This was followed by a further 40 cycles refining against data to 3.1 Å, resulting in a final free R-factor of 0.372 and R-factor of 0.363 (for data in the range 15.0 - 3.1 Å).

*3.5.7.2 X-PLOR simulated annealing refinement.*

Refinement by simulated annealing was next conducted using X-PLOR (Brunger et al., 1987). Molecular dynamics is used to simulate the motion of the protein atoms, yielding a cluster of energetically allowed structures for a given temperature. The simulated annealing protocol consists of raising the temperature to a sufficiently high level to overcome energy barriers and then slowly cooling the system to its energy minimum. The crystallographic residual and the geometric restraints are expressed as energy terms in the minimization.

NCS restraints were imposed on the refinement. An overall temperature factor was refined and no contribution for the scattering of the bulk solvent was included. The model was restrained to the Engh and Huber geometry library (Engh and Huber, 1991). After several rounds of simulated annealing refinement and model rebuilding the R-factor had dropped to 24.7% and the free R-factor to 30.7% for all reflections between 10.0 and 3.1 Å. The rms deviations from ideal geometry for the model at this stage were 0.013 Å in bond lengths and 2.6° in bond angles.

*3.5.7.3 TNT restrained least squares refinement*

The TNT restrained least squares package (Tronrud et al., 1987) was used for least-squares restrained refinement of the model. This program was used as it has an elegant and easy method of implementing NCS constraints during the refinement and also allows the simultaneous refinement of atomic positions and temperature factors. The NCS restraints on the Gower II hemoglobin tetramer were replaced with NCS constraints and one temperature factor per residue was refined. The resolution limits were increased to include the higher resolution data, and because a bulk solvent correction was included in the scaling, the lower resolution terms were also used (20-2.9 Å). After several rounds of refinement and rebuilding, and the subsequent introduction of restrained individual isotropic temperature fac-

tors into the refinement the crystallographic R-factor and free R-factor were 0.213 and 0.273 respectively.

*3.5.7.4 TNT maximum likelihood refinement*

The recent implementation of a maximum likelihood target in TNT (Pannu and Read, 1996) allowed this method to be utilised in the final stages of the Gower II hemoglobin refinement. The function minimised in least squares amplitude refinement is:

$$\sum_{hkl} w\,(|F_o| - k|F_c|)^2 \qquad \textit{Eq. 3.6}$$

where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, $k$ is a scale factor, and $w$ a weight. The assumption made in refining against this residual is that the deviation between $|F_o|$ and $|F_c|$ is a Gaussian with mean zero and a standard deviation that is independent of the model parameters. However this is not appropriate, as phase errors in the calculated structure factors, which can be large when the model is incomplete or contains many errors, are not taken into account. A maximum likelihood treatment has been formulated (Read, 1990; Bricogne, 1993) that is more suited for protein structure refinement and has been implemented into various programs (Pannu and Read, 1996; Adams et al., 1997; Murshudov et al., 1997). This method is based on determining the probability of making a measurement, given the model and its errors (misplaced and missing atoms), as well as errors in the intensity measurements. The effects of these model errors on the calculated structure factors are quantified by calculating $\sigma_A$ values (Read, 1986). The free-R set of reflections is used to calculate $\sigma_A$ to prevent overfitting of the amplitudes and hence underestimating the errors in the calculated structure factors. Test calculations using maximum likelihood refinement on a number of protein structures have given dramatic results, yielding models with twice the improvement in average phase error, with clearer and less biased electron density maps in comparison to a least squares minimization (see references above).

The Gower II hemoglobin model was well refined after the least squares refinement (Rf=0.273, R=0.213). However cycles of maximum likelihood refinement, in conjunction with model building, did result in improvements in the model. The free R-factor decreased to 0.269 and the R-factor increased to 0.225. The actual decrease in the free R-factor was greater than these statistics suggest as scale factors are

refined for both the working and free sets in the TNT least squares refinement, whereas the maximum likelihood implementation refines the scale factor against all the reflections. The additional scale factor typically resulted in the free R-factor being underestimated by around 0.5% in the least squares case.

*3.5.7.5 X-PLOR least squares refinement*

A final round of refinement was conducted in X-PLOR to incorporate the mask-type bulk solvent correction unique to this program. Tests calculations have shown this method to give an improved description of the solvent scattering compared to the Babinet type model (Section 3.5.5). The crystallographic R-factor at the end of refinement was 0.185, with a free R-factor of 0.232. The progress of the Gower II hemoglobin refinement is summarised in Table 3.15.

**Table 3.15** Overview of Gower II Hb refinement

| Method | Program | Data (Å) | NCS Model | Bulk Solvent Model | B factor Model | Final R free | Final R |
|---|---|---|---|---|---|---|---|
| RBR | X-PLOR | 10.0-3.1 | none | None | Overall | 0.372 | 0.363 |
| SAR | X-PLOR | 10.0-3.1 | restrained | None | Overall | 0.307 | 0.247 |
| LSQR | TNT | 20.0-2.9 | constrained | Babinet | Isotropic | 0.273 | 0.213 |
| MLR | TNT | 20.0-2.9 | constrained | Babinet | Isotropic | 0.269 | 0.225 |
| LSQR | X-PLOR | 40.0-2.9 | restrained | Mask | Grouped | 0.232 | 0.185 |

Abbreviations: RBR=rigid-body refinement, SAR=simulated annealing refinement, LSQR=least squares refinement, MLR=maximum likelihood refinement.

### *3.5.8 Quality of the final model*

The final model comprises residues 1-141 of the alpha subunits, i.e. the complete polypeptide chain, and residues 1-145 of the epsilon subunits, having the C-terminal residue (His 146) omitted with no interpretable electron density. A number of side-chains, predominately on the surface of the molecule, with little or no electron density present, have not been included in the model (six in $\alpha_1$, seven in $\alpha_2$, twelve in $\varepsilon_1$ and eleven in $\varepsilon_2$, Table 3.16). These residues have been modelled as alanine.

**Table 3.16** Side-chains not included in the Gower II hemoglobin atomic model

| $\alpha_1$ | $\alpha_2$ | $\varepsilon_1$ | $\varepsilon_2$ |
|---|---|---|---|
| 23 Glu | 23 Glu | 2 His | 2 His |
| 50 His | 30 Glu | 22 Glu | 6 Glu |
| 56 Lys | 50 His | 59 Lys | 22 Glu |
| 60 Lys | 56 Lys | 61 Lys | 76 Lys |
| 78 Asn | 60 Lys | 65 Lys | 80 Asn |
| 141 Arg | 78 Asn | 76 Lys | 82 Lys |
| | 141 Arg | 82 Lys | 87 Lys |
| | | 87 Lys | 95 Lys |
| | | 90 Glu | 120 Lys |
| | | 95 Lys | 143 His |
| | | 120 Lys | 144 Lys |
| | | 121 Glu | |
| | | 143 His | |

The free R-factor for the final Gower II hemoglobin model is 0.232 (1485 reflections), with a conventional crystallographic R-factor of 0.185 (13113 reflections), for all data in the range 40.0-2.9 Å. The model is tightly restrained, with rms deviations from ideal bond lengths of 0.0064 Å and ideal bond angles of 1.16°. The rms difference (calculated on all Cα atoms) between the two α chains, which is a measure of the deviation from true NCS for the two subunits, is 0.015 Å. The rms difference for the ε chains, calculated in the same manner is 0.048 Å. The reason for the greater divergence between the two ε chains is due to a shortcoming in the method in which X-PLOR implements NCS restraints. Residues that are not identical, i.e. a lysine that may be modelled in one monomer but is disordered in the other and is included as an alanine, cannot be restrained, even those atoms that are common (e.g. main-chain and Cβ atoms). The ε chain has non-equivalent residues at a greater number of positions (12) than the α (3) and hence displays a larger NCS difference between its subunits.

The main-chain torsion angles ($\phi$ and $\psi$) were not restrained during the refinement (though Ramachandran plots were referred to during model rebuilding) and hence

the distribution of these values can be used as a method of validating the structure. A Ramachandran plot of main-chain torsion angles calculated with PROCHECK (Laskowski *et al.*, 1993) shows that 92.1% of residues lie in the most favoured region, with none in the disallowed area (Figure 3.9).

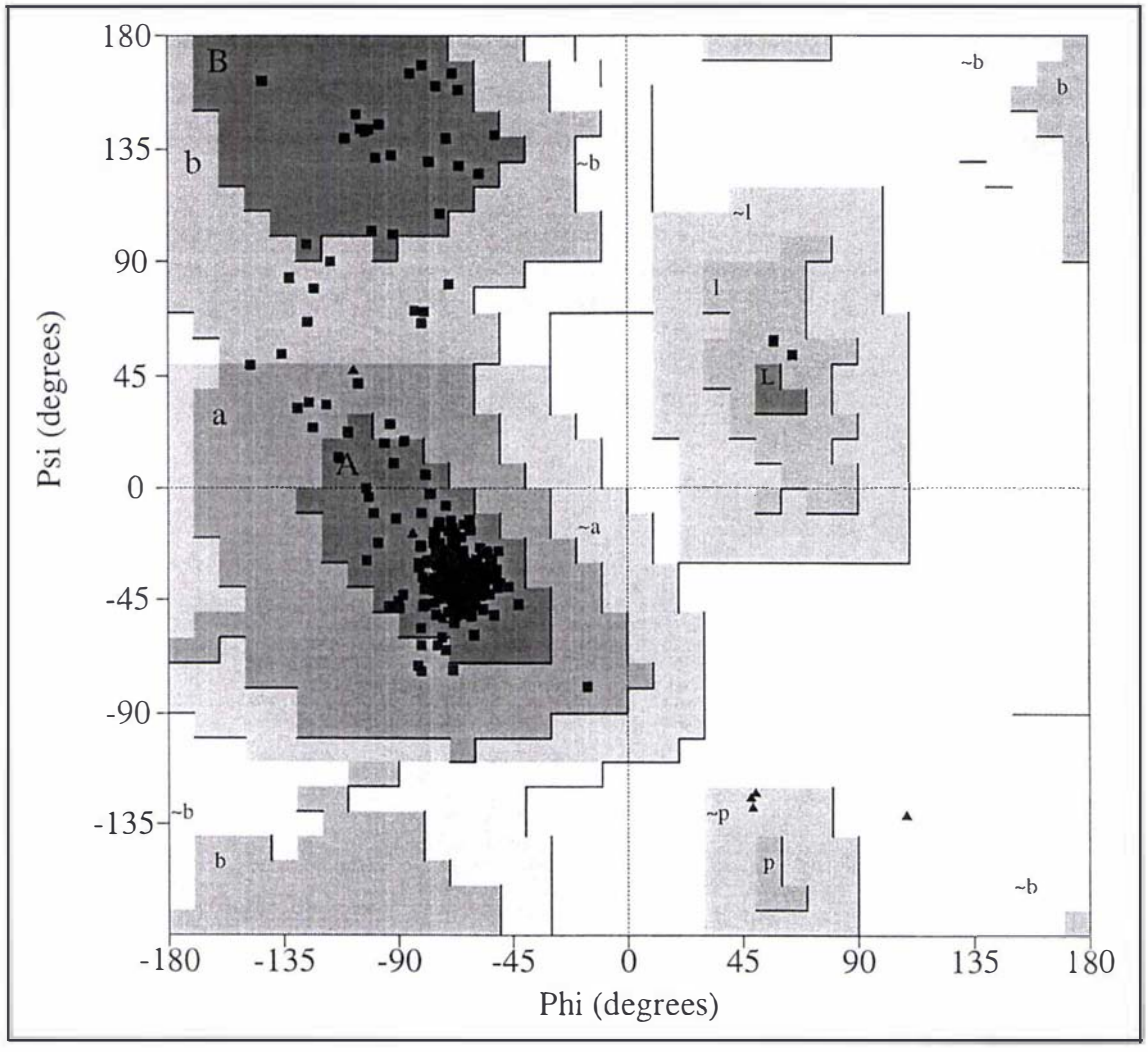



**Figure 3.9** Ramachandran plot for the refined Gower II Hb tetramer

Glycine residues are represented as triangles, with all other residues shown as squares. Figure produced from PROCHECK (Laskowski et al., 1993)

A measure of how well the atomic model fits the final electron density map can be expressed by calculating real-space correlation coefficients (cc) for each residue of the polypeptide chain. The coefficients calculated with CNS (Brunger et al., 1997) for the final model as a function of residue number for each subunit of the Gower II hemoglobin tetramer are shown in Figure 3.10.

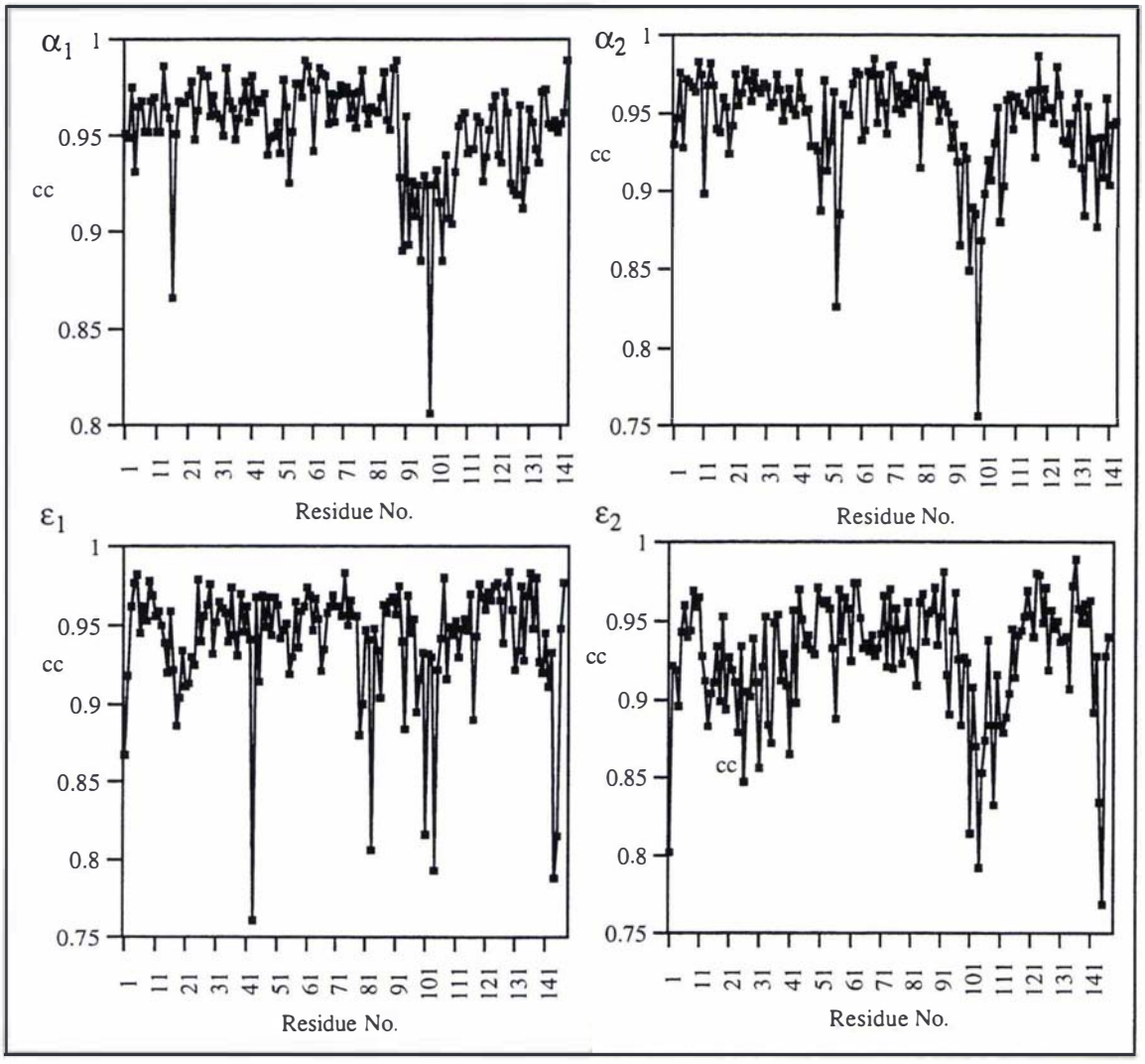



**Figure 3.10** Real-space correlation coefficients for the Gower II Hb subunits

The heme group and carbon monoxide ligand have been included as the two C-terminal residues in each monomer.

The outliers in the real-space correlation coefficient plots can be predominantly attributed to charged residues on the surface of the molecule with poor side-chain density, for example Asp $\varepsilon_1$43: Glu $\varepsilon_1$101, $\varepsilon_2$101: Lys $\alpha_1$99, $\alpha_2$99, $\varepsilon_1$104 and the $\varepsilon$ C-termini. It is interesting to note that residues 91-103 of both $\alpha$ chains, a region of the sequence that corresponds to the FG loop and the N-terminal part of the G helix is described as fitting the density less well, an effect not obvious from visual inspection of the atomic model and maps. This region (FG corner) is part of the structure that participates in the allosteric switch region during the R to T transition.

The average temperature factors for the Gower II hemoglobin tetramer main-chain and side-chain atoms are 49 Å$^2$ and 50 Å$^2$ respectively, compared to a calculated Wilson B of 62 Å$^2$. A more detailed analysis of the temperature factors is given in Table 3.17. It is interesting to note that the average temperature factors are higher for the epsilon chains that for the alpha subunits. This may reflect an increased atomic motion within the epsilon chains, but could also give an indication of the failure of the refinement to completely converge at this resolution; the model for the alpha chain was essentially unchanged from that found for the adult hemoglobin, which had been refined at 1.7 Å yielding a precise and accurate model. It is also worth noting that when a protein structure is refined with NCS constraints or very tight restraints the temperature factors can incorporate any breakdown of the NCS and hence will be inflated.

**Table 3.17** Average temperature factors (Å$^2$) for the Gower II hemoglobin

| Atoms | Gower II Hb subunit | | | |
|---|---|---|---|---|
| | $\alpha_1$ | $\alpha_2$ | $\varepsilon_1$ | $\varepsilon_2$ |
| Main-chain | 39 | 44 | 58 | 59 |
| Side-chain | 42 | 44 | 57 | 61 |
| Heme | 37 | 39 | 78 | 60 |
| Ligand | 21 | 39 | 41 | 43 |
| All | 40 | 44 | 58 | 60 |

The four heme groups are well defined, except for the propionic acid groups on the ε hemes, which have weak density. Clear density is observed for a CO ligand bound on each heme. The SIGMAA coordinate error (Read, 1986) is 0.35 Å (calculated with all reflections). No water molecules are included in the model owing to the low observation to parameter ratio at this moderate resolution. A summary of the final model and refinement statistics is included in Table 3.18.

**Table 3.18** Refinement and model statistics

| | |
|---|---|
| R-factor (all data 40.0-2.90 Å) | 0.185 (0.301) |
| Free R-factor (all data 40.0-2.90 Å) | 0.232 (0.366) |
| Bond length rms deviation | 0.0064 Å |
| Bond angle rms deviation | 1.16° |
| NCS Alpha subunit rms difference | 0.015 Å |
| NCS Epsilon subunit rms difference | 0.048 Å |
| G factor[#] | 0.5 |
| Residues in most favoured region of Ramachandran plot[#] | 92.1% |
| Residues in disallowed region of Ramachandran plot[#] | 0% |
| Average main-chain temperature factor | 49 $Å^2$ |
| Average side-chain temperature factor | 50 $Å^2$ |
| Number of protein atoms in model | 4234 |
| Number of heme atoms in model | 172 |
| Number of CO atoms in model | 8 |

The R-factors given in parentheses refer to the outermost 2.97-2.9 Å shell.

[#]As defined by PROCHECK (Laskowski et al., 1993)

Chapter 4

# Gower II Embryonic Hemoglobin: Structure and Function

## 4.1 QUATERNARY STRUCTURE

The Gower II hemoglobin molecule assembles into the well characterised hemoglobin tetramer of two hetero-dimers (Figure 4.1 and Figure 4.2). Each dimer comprises the shorter 141 residue α polypeptide and the longer 146 amino acid ε chain. Each chain binds a heme group via a proximal histidine.

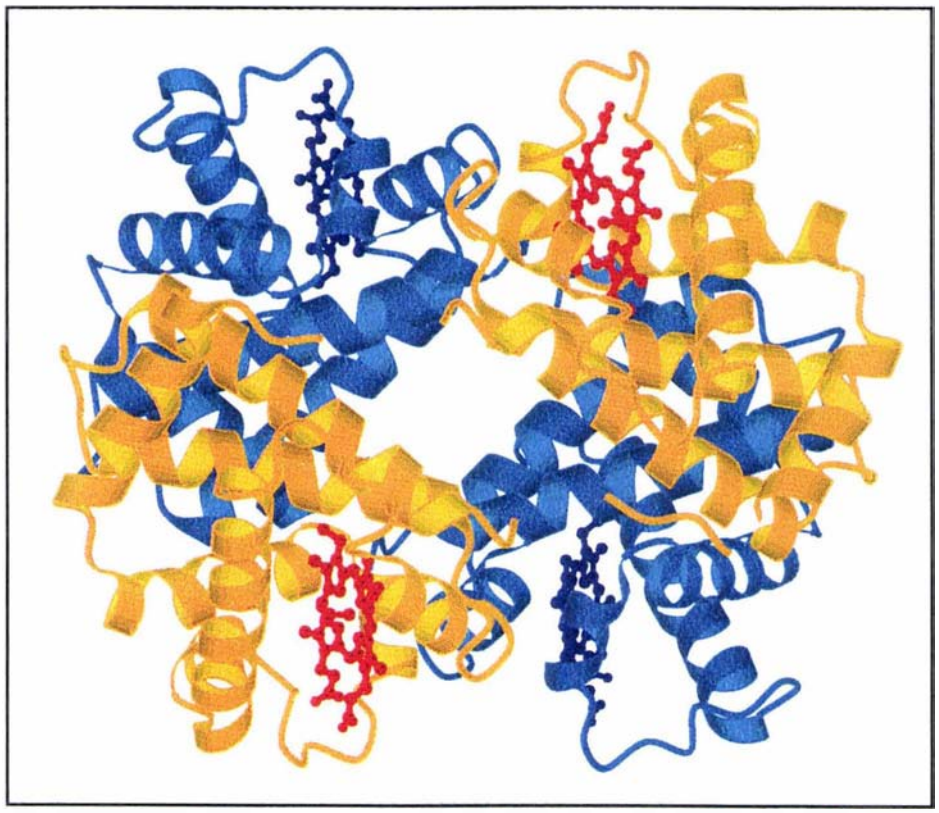


**Figure 4.1** Ribbon diagram of the Gower II embryonic hemoglobin tetramer

The epsilon chain is represented in gold and the alpha chain shaded blue. The heme groups and carbon monoxide ligands are shown as a ball and stick representation. The figure was prepared with MOLSCRIPT (Kraulis, 1991).

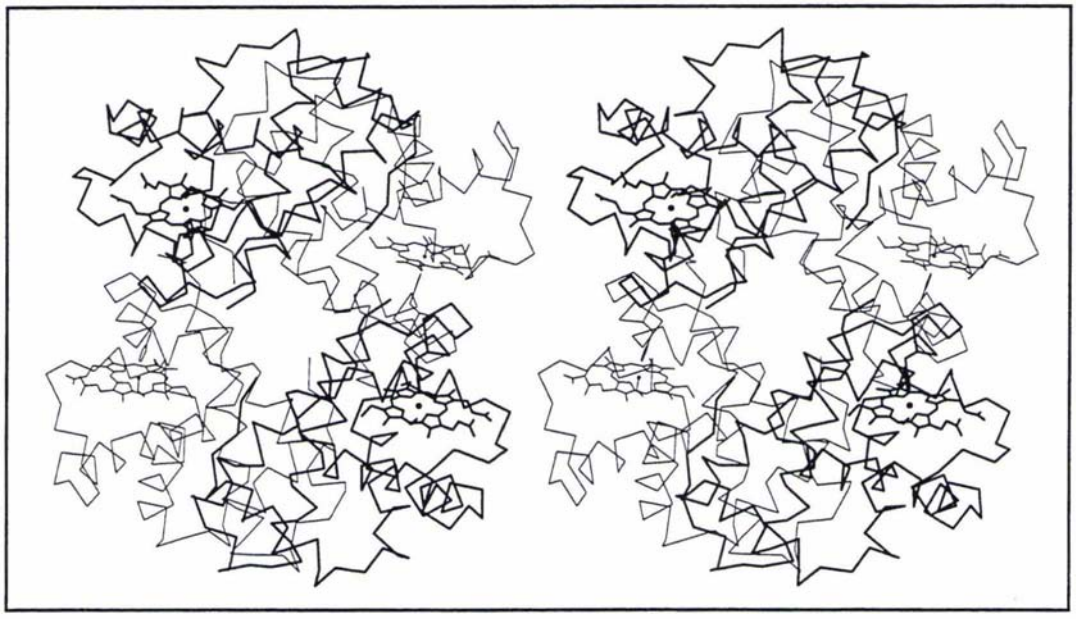


**Figure 4.2** Stereo Cα trace of the Gower II embryonic hemoglobin tetramer

The epsilon chain is shown in thicker lines. The figure was prepared with MOL-SCRIPT (Kraulis, 1991).

Two quaternary states for the liganded hemoglobin tetramer have been characterised, the R state (Shaanan, 1983) and the R2 (Silva et al., 1992) or Y state (Smith et al., 1991; Smith and Simmons, 1994). The quaternary structure of carbonmonoxy embryonic Gower II hemoglobin lies between that of the R2 and R states, though closer in orientation to the R2 tetramer. The rotation required to overlay the Gower II αε and adult αβ dimers, after superposition on the other half of the tetramer, is 3° for the R2 state and 9° for the R state (Figure 4.3). The Gower II hemoglobin, like the R2 or Y adult hemoglobin quaternary structures was crystallized from low salt conditions, in contrast to the adult R state model which was crystallized from 2.5M sodium/potassium phosphate (Shaanan, 1983). It appears that crystallization of the R2 state is favoured by low salt conditions (Silva et al., 1992; Srinivasan and Rose, 1994; Smith and Simmons, 1994). Srinivasan and Rose suggest that the R state is an intermediate quaternary structure trapped between the R2 and T states due to the high salt crystallization conditions (Srinivasan and Rose, 1994), an idea supported by the studies of cross-linked hemoglobins (Schumacher et al., 1997). Other analyses have lead to the proposal that the R2 state may be an intermediate on the R to T pathway, (Silva et al., 1992) or a modified R state (Janin and Wodak, 1993). The rms differences for all main-chain atoms between the Gower II

hemoglobin, and adult R2 (PDB code 1BBB) and R (PDB code 1HHO) state tetramers are 0.8 Å and 1.6 Å respectively. Superpositions calculated with all the main-chain atoms for just one $\alpha_1\varepsilon_1/\alpha_1\beta_1$ dimer yield rms differences of 0.7 Å and 1.1 Å for the R2 and R states. By comparison the rms difference between the Gower II and deoxy adult T state tetramers (Fermi et al., 1984, PDB code 2HHB) is 3.0 Å, and 1.1 Å for superposition of the dimers.

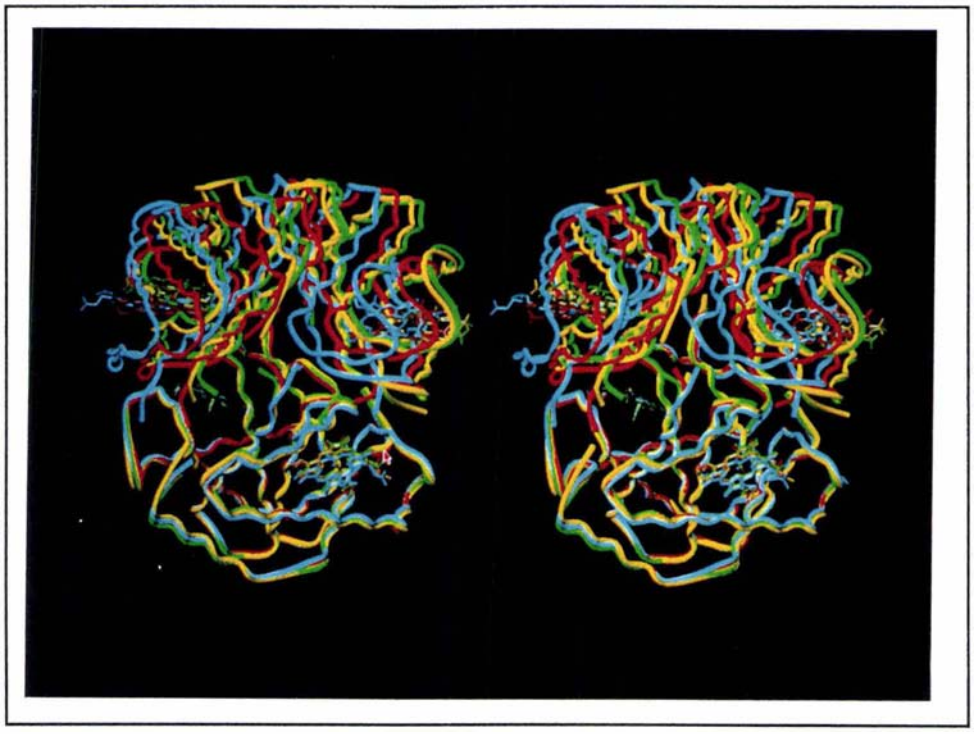


**Figure 4.3** Stereo Cα trace of the Gower II (yellow), adult R2 state (green), adult R state (red) and adult T state (blue) Hb tetramers.

The superposition was calculated on one αβ dimer. The figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

## 4.2 ALPHA SUBUNIT

### *4.2.1 Tertiary structure*

The α subunit is strictly conserved between the human adult ($\alpha_2\beta_2$), fetal ($\alpha_2\gamma_2$) and Gower II embryonic ($\alpha_2\varepsilon_2$) hemoglobins. The tertiary structure of the Gower II alpha subunit is essentially identical to that of the alpha subunit found in the other members of the human hemoglobin family (Figure 4.6, on page 75). The rms difference of all main-chain atoms of the embryonic alpha chain is 0.3 Å when compared to the

adult α polypeptide, a value close to the estimated error of the coordinates at this resolution. The largest conformational change observed is the displacement of the C-terminal region of the loop between the C and E helices (the CD loop in standard hemoglobin nomenclature, corresponding to residues 47-52) resulting in a shift of 1.2 Å of the Cα atom of residue 50 away from the central cavity on the tetramer dyad axis. Flexibility in this region is indicated by high temperature factors for this part of the structure (in the region of 80 $Å^2$) and the lack of any side-chain density for the side-chain of histidine 50. This loop shifts as a rigid body maintaining a similar conformation to that found in the adult, with main-chain torsion angles characteristic of R and R2 state hemoglobins (Srinivasan and Rose, 1994). The embryonic Gower II C-terminal dipeptide is found in the more extended R2 conformation as compared to the R state. There is a shift of residue 141 towards the tetramer twofold axis relative to the adult R2 α subunit (Cα movement of 1.3 Å). This region is highly solvated in both the adult liganded quaternary structures (Silva et al., 1992; Shaanan, 1983) and this also appears to be the case in the Gower II structure, with no side-chain density visible for the C-terminal arginine side-chain.

### *4.2.2 Heme pocket*

The structure of the α subunit heme pocket in Gower II hemoglobin is extremely similar when compared to the corresponding subunit of adult R and R2 hemoglobin. A superposition of the α heme pocket (all atoms of helices E and F, and heme group) results in an rms difference of 0.5 Å from the adult R2 molecule. Though the present resolution is limited, the refinement and the electron density consistently favour an angular orientation for the carbon monoxide bound to the heme (Figure 4.4 and Figure 4.5).

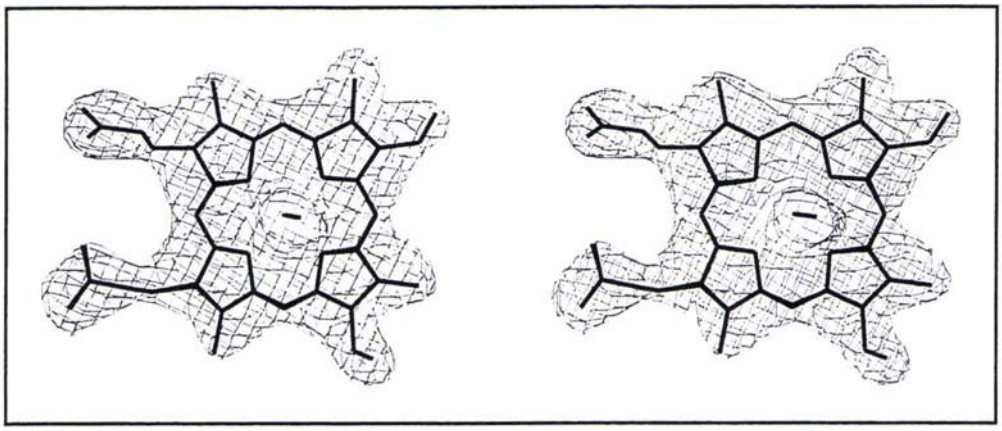


**Figure 4.4** $2m|F_o| - D|F_c|$ map for the $\alpha_1$ heme.

The map is contoured at 1 sigma. The figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

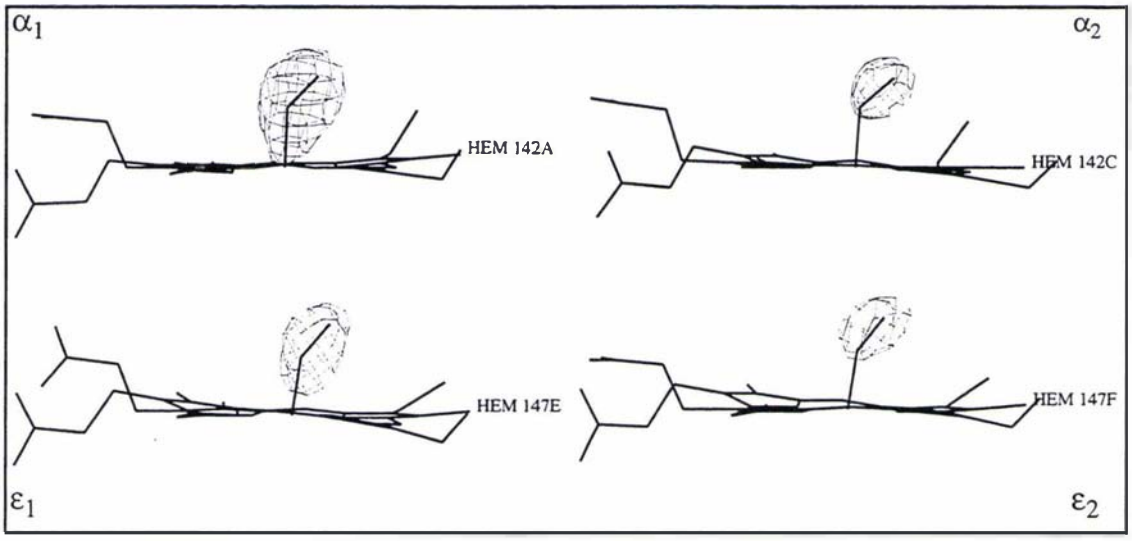



**Figure 4.5** $m|F_o| - D|F_c|$ omit maps for the ligand at each of the Gower II Hb heme groups.

The omit map is contoured at 5 sigma in each case. The figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

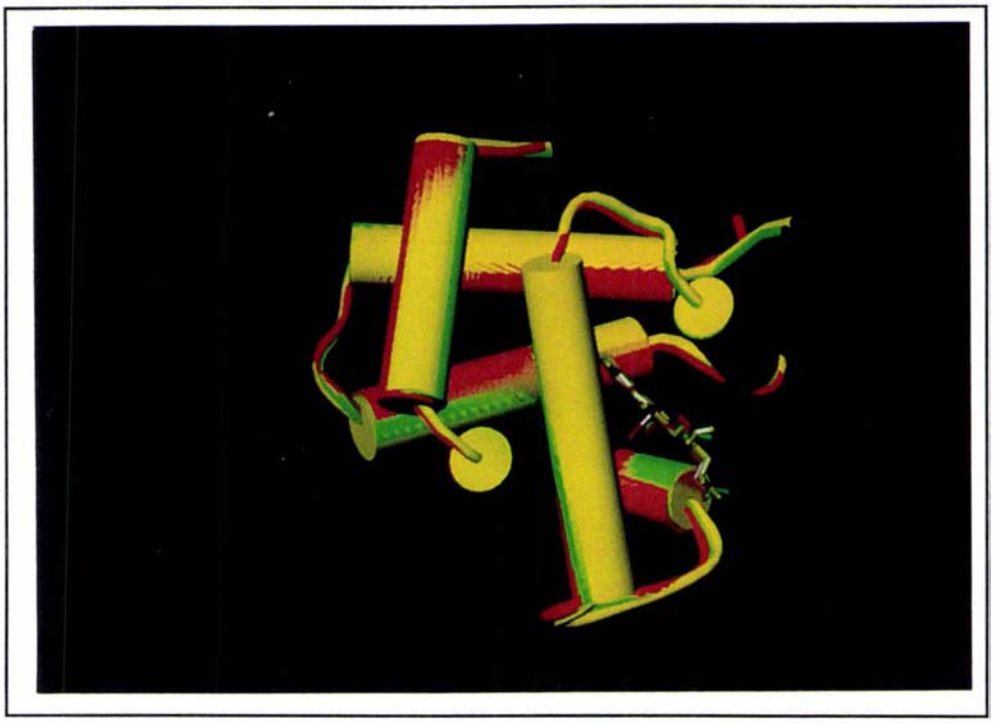


**Figure 4.6** Superposition of the Gower II α (yellow), adult R2 state α (green) and adult R state α (red) Hb subunits (calculated on all Cα atoms)

These figures were prepared with TURBO-FRODO (Roussel and Cambillau, 1991)

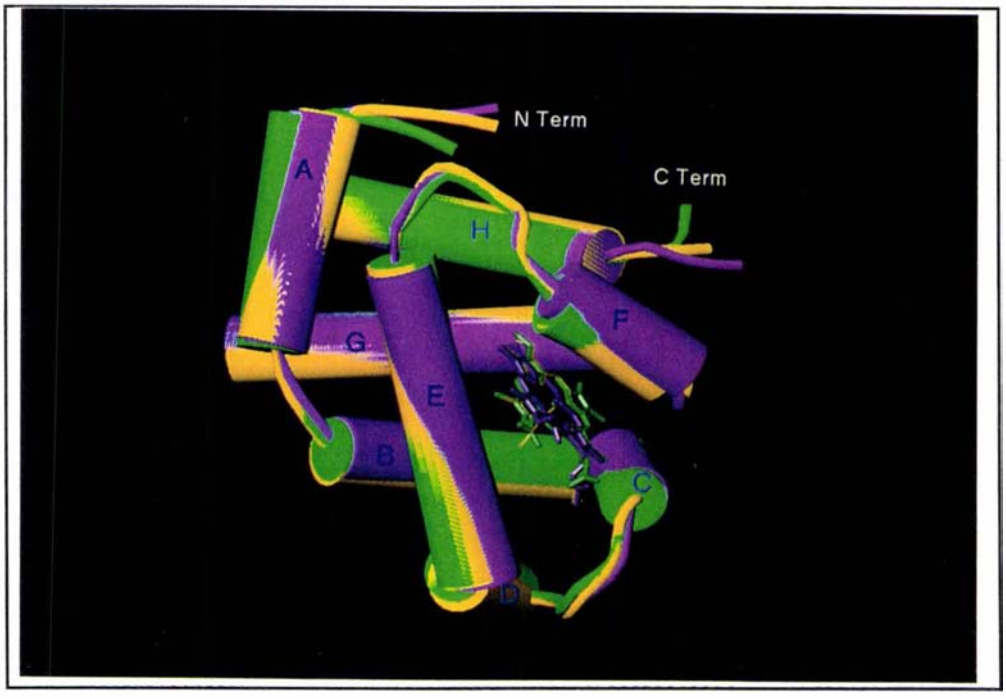



**Figure 4.7** Superposition of the Gower II ε (yellow), adult R2 state β (green) and fetal γ (purple) Hb subunits (calculated on BGH frame)

## 4.3 EPSILON SUBUNIT

### *4.3.1 Tertiary structure*

The ε subunit found in the Gower I ($\zeta_2\varepsilon_2$) and Gower II ($\alpha_2\varepsilon_2$) human embryonic hemoglobins comprises a polypeptide chain of 146 amino acids (identical in length to the β and γ subunits) and differs from the adult β chain at 36 positions (75% sequence identity), and from the γ chain, found in fetal ($\alpha_2\gamma_2$) and embryonic Portland ($\zeta_2\gamma_2$) hemoglobins at 30 positions (79% sequence identity). The ε subunit exhibits the classic globin fold, having a very similar conformation to both the β and γ chains. Baldwin and Chothia (Baldwin and Chothia, 1979) established that the regions of the hemoglobin tetramer that move the least during the T to R transition are those secondary structure elements at the $\alpha_1\beta_1$ interface, namely the B, G and H helices, and proposed that this be the most suitable reference frame for the superposition of hemoglobin structures for comparative analysis (BGH frame). Superposition on these helices shows that the overall tertiary fold of the ε chain is very similar to that of the β and γ subunits (Figure 4.7). The rms differences for the position of all main-chain atoms of the ε subunit when superimposed on the β and γ chains are 0.7 Å and 0.8 Å respectively, whereas the corresponding rms differences for superposition on the BGH helices only are 0.5 Å and 0.6 Å.

### *4.3.2 The amino terminus and A helix*

The most significant difference between the ε and β subunits is a shift of the N-terminus and A helix, with a similar change also observed in the T state γ chain of fetal hemoglobin (Frier and Perutz, 1977). The A helix moves towards the central cavity of the tetramer in the direction of the R to T transition. The N-terminal region of the subunit contains a large number of amino acid substitutions relative to the whole sequence. The shift in the A helix appears to be due to the increased steric bulk of the phenylalanine side-chain (β3 Leu → ε3 Phe) packing against the EF loop, close to the substitution of another large side-chain; β78 Leu → ε78 Met. Flexibility in the N-terminus is aided by the replacement β5 Pro → ε5 Ala. The substitution β130 Tyr→ ε130 Trp results in the loss of a hydrogen bond between the β130 Tyr OH and the carbonyl oxygen of β11 Val, which would hold the A helix closer to the H helix in the adult β subunit. The extra size of the ε130 Trp side-chain causes the adjacent ε15 Trp side-chain to swing closer to the E helix, where a β72 Ser → ε72 Gly substitution has made room for the ε15 Trp to pack against the helix. In the β chain the 15 Trp Nε1 and the 72 Ser Oγ are hydrogen bonded. The conformational changes in this N-terminal region and A helix are stabilized by an additional hydrogen bond between

the carbonyl oxygen of His 2 and Nζ of Lys 132 (2.8 Å) compared to a very much longer distance between these two atoms of 3.8 Å in the R2 β chain, with no interaction at all in the R state β model due to a peptide flip at this position. Variability in the relative position of the A helix has been previously reported in the crystal structures of the "beta-like" chain of hemoglobins from other mammals (Katz et al., 1994; Perutz et al., 1993).

### *4.3.3 Internal regions*

There are a number of substitutions between the ε and β chains that are located within the internal core of the molecule. A large number of these occur in the pocket formed by the A, B, E, G and H helices, located away from the heme pocket. The substitution at residue 130, β Tyr → ε Trp results in the loss of a hydrogen bond with the carbonyl oxygen of ε11 and also a large shift of Trp 15, coupled with a smaller rotation of the A helix to relieve steric clashes between the side-chains. This movement allows more room for the larger Met side-chain ε18 (Val in β). Apart from these changes little other structural difference is observed. The majority of other internal amino acid substitutions are conservative.

### *4.3.4 Heme pocket*

Superposition of the E helices from the embryonic ε chain, adult R β and R2 β chains reveals no change in the relative positions of the E and F helices, and heme group. There is a small shift in the position of the EF loop (residues 77-84, between the E and F helices which form the heme pocket). This change in conformation of the EF loop does not affect the positions of the E and F helices. Though the resolution of the data is limited, the geometry of the liganded epsilon heme group appears very similar to that of the liganded beta heme, exhibiting a comparable puckering. A superposition of the heme pocket (all atoms of helices E and F, and heme group) results in an rms difference of 0.7 Å from the adult R2 β chain. The carbon monoxide ligand again appears to be bound in an angular mode (Figure 4.8).

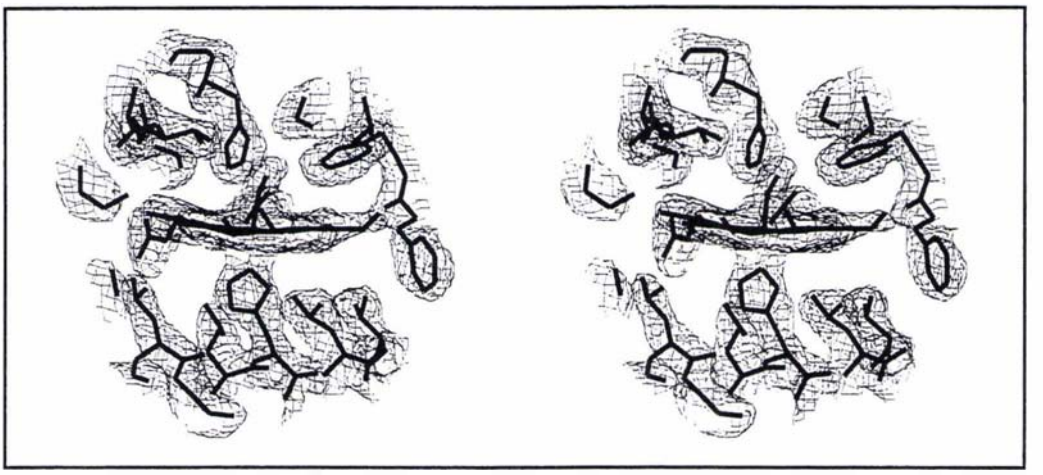


**Figure 4.8** $2m|F_o| - D|F_c|$ map of the ε heme pocket

The map is contoured at 1 sigma. The figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

There is a single amino acid substitution contained within the heme pocket on the E helix at position 70β Ala → ε Ser. The serine Oγ hydrogen bonds back to the main-chain carbonyl oxygen of Val 67, as well as packing against the methyl substituent on the B pyrrole ring of the heme. This substitution also occurs in some other mammalian hemoglobins (cow, pig and horse).

One substitution with important implications for the T to R transition in the Gower II hemoglobin is 83β Gly→ ε Pro, in a position at the start of the F helix. Relative to the adult molecule the Gower II hemoglobin has a less stable T state (Brittain et al., 1997). The introduction of a rigid proline ring would be expected to decrease the flexibility of the EF loop and restrict the manner in which the F helix can 'react' to changes in ligation at the heme, increasing the steric strain of the Gower II T state. This substitution may also be partially responsible for the twenty five-fold increase in heme binding of the ε subunit relative to the adult β chain found for the respective methemoglobins (Robson and Brittain, 1996). In the ε chain Lys 66 appears to form a weak ionic interaction with the heme propionate, with the Nζ of the Lys 3.6 Å from the carboxyl oxygen (Figure 4.9). The Lys 66 residue is conserved amongst the human hemoglobin family, but only in the fetal γ subunit structure is the side-chain in a similar conformation to that found in the ε chain. This may also explain the higher heme affinity of the γ subunit (ten times greater than the adult β) measured in the Portland ($\zeta_2\gamma_2$) embryonic methemoglobin.

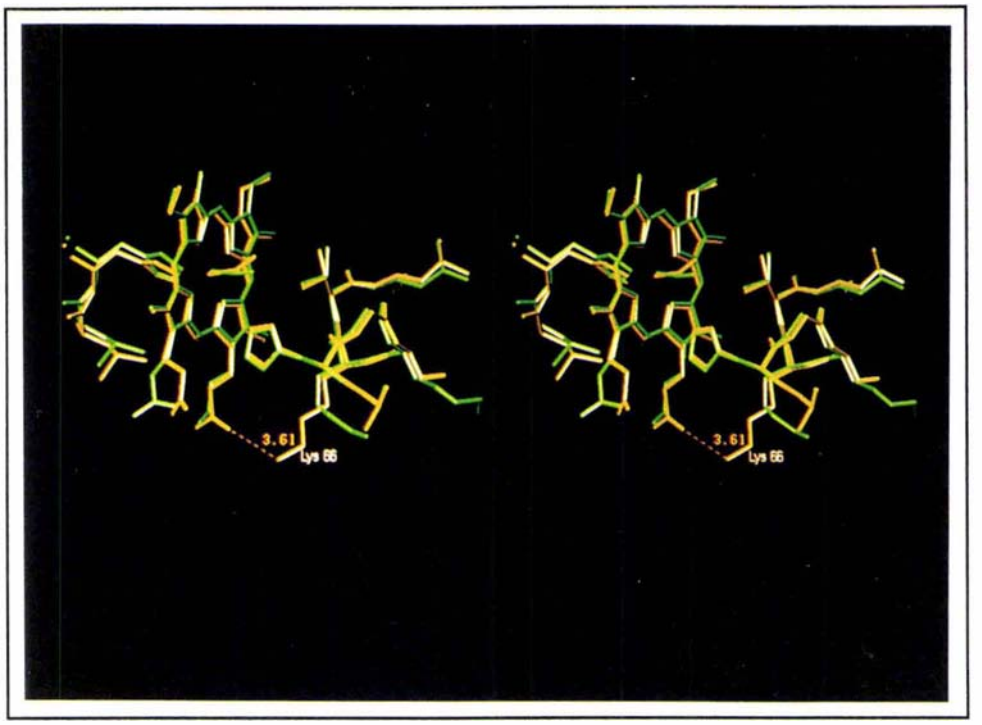



**Figure 4.9** Stereo view of the Gower II ε (yellow) and adult R2 state β (green) heme groups.

The superposition was calculated on the E and F helices and the heme group. This figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991)

### *4.3.5 The C-terminus and reactive Cys 93*

The β subunit C-terminus, and Cys 93 situated on the FG corner, is one of the regions that undergoes major change during the allosteric switch between the R and T quaternary states. In the T state a number of hydrogen bonding interactions are formed by the C-terminal residues: β146 His forms two salt bridges, one involving the C-terminus across the $\alpha_1\beta_1$ interface to α40 Lys, and a side-chain interaction with β94 Asp, while β145 Tyr interacts with the carboxyl group of β98 Val (for a detailed description see Fermi, 1975; Baldwin and Chothia, 1979). In the liganded forms of the molecule the salt bridges involving the terminal residues described above are broken, and these amino acids disordered. This increased flexibility at the C-terminus on ligand binding results in β145 Tyr alternating between a major and minor site in conjunction with β93 Cys (Baldwin, 1980; Shaanan, 1983). For the deoxyhemoglobin β chain Cys 93 adopts the trans rotamer with the side-chain pointing towards the outside of the molecule, and Tyr 145 fills a pocket between the F, G and H helices. For the liganded quaternary R or R2 state this pocket is filled by the Cys 93 side-chain (now in the gauche minus rotamer), with the Tyr now shifted to

the surface. The electron density maps for oxyhemoglobin suggested that Tyr 145 and Cys 93 were alternating between the major liganded conformation (with occupancies of 0.8 and 0.7 respectively) and the minor deoxyhemoglobin-like conformation (with occupancies of 0.2 and 0.3 respectively, Shaanan, 1983). In the liganded methemoglobin the equilibrium is shifted towards the deoxyhemoglobin conformation with the Tyr filling the internal pocket and Cys 93 being on the surface (Heidner et al., 1976).

The conformation of the Cys 93 side-chain in the Gower II hemoglobin ε chain is similar to that found for deoxyhemoglobin (i.e. the trans rotamer). The electron density indicates unambiguously that the sulphydryl is clearly in this position, with no density present for an internal location in the pocket of the F, G and H helices for this side-chain. The C-terminus of the ε chain is disordered, as is common for liganded hemoglobins, with no interpretable electron density present for His 146. A peak is visible in the maps for the side-chain of Tyr 145 in the FGH pocket, though the temperature factors are very high for this residue (main-chain ~139 $Å^2$, side-chain ~133 $Å^2$) and no density is visible for the main-chain atoms. A large positive peak (over 4σ) is visible at the end of the Cys side-chain, 2.2 Å from the Sγ, suggesting that some species may be bound at this site (Figure 4.10). It is known that certain groups can bind to this cysteine, hence the 'reactive' name commonly attributed to it. The cysteine can react with mercury compounds (Perutz et al., 1960) and nitric oxide (Gow and Stamler, 1998). Hence the possibility that dithionite (or some derivative of this molecule) could bind at this site must be considered, resulting in Cys 93 being held in the trans conformation.

The Gower II hemoglobin was crystallized in the presence of 2mM sodium dithionite in order to limit oxidation of the heme iron. It may be the case that dithionite has bound to the Cys 93 on the outside of the molecule, holding it in the conformation normally found for met- and deoxyhemoglobin.

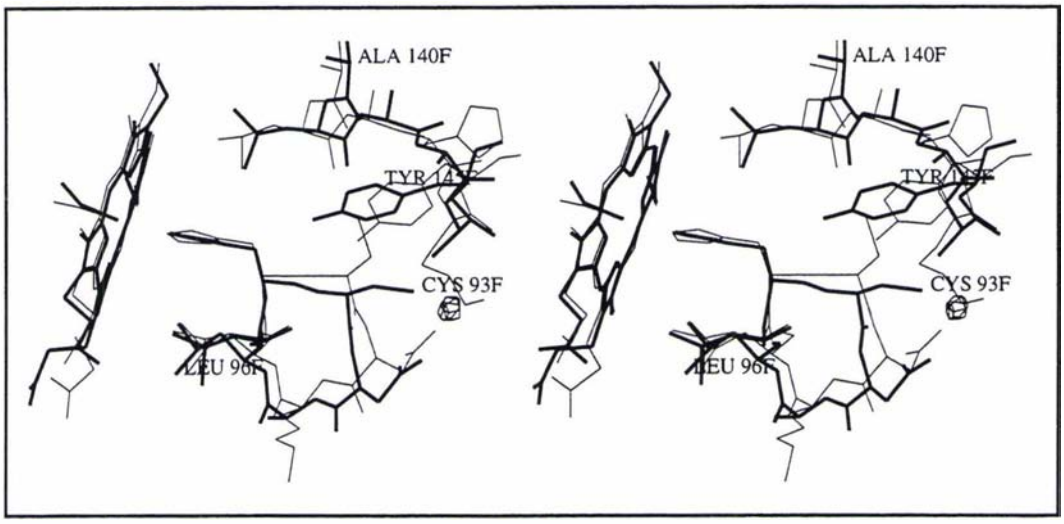



**Figure 4.10** The Gower II ε (thick) and adult R2 state β (thin) Cys 93 and C-termini.

The $m|F_o| - D|F_c|$ map is contoured at 4 sigma (The coordinate sets were superimposed on the E helix and heme group). This figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

The alternative hypothesis is that despite conducting the crystallization experiments under an atmosphere of carbon monoxide, oxidation of the Gower II hemoglobin did occur. This may have happened during mounting of the crystal, which was conducted in air. Estimation of the met content in hemoglobin crystals is problematic. A common practice is to record a UV/visible spectrum on a dissolved crystal (Liddington et al., 1992; Katz et al., 1994), though it has been suggested that a large proportion of any measured oxidation occurs during the process of the crystal dissolving (Paoli et al., 1996). The method used by Paoli and co-workers to confirm the identity of the ligand was an examination of the ligand temperature factors, as well as inspection of the electron density for the ligand, both with the ligand present and with it omitted for the map calculations. Furthermore the structure was refined assuming the presence of a water ligand, and the difference maps calculated from this model were inspected for evidence of a diatomic ligand. This protocol was followed for the Gower II hemoglobin model. The electron density for the ligands appears consistent for a diatomic carbon monoxide ligand, both in a $2m|F_o| - D|F_c|$ map in which the ligand was included in the map calculation (Figure 4.8, on page 78), and a $m|F_o| - D|F_c|$ omit map (Figure 4.5, on page 74). Analysis of the temperature factors for the ligand (see Table 3.17, on page 68) is complicated by the manner in which these were treated in the refinement. A single temperature factor was refined for the heme group, resulting in the iron B factor becoming inflated as a result of disorder in the

external parts of the heme, principally the propionates. A single temperature factor was refined for the two atoms of the carbon monoxide ligand, with these values appearing consistent with a fully occupied diatomic ligand. For refinement with a single oxygen ligand, as for aquo-methemoglobin, the free R-factor increased slightly from 0.232 to 0.235, with the conventional R-factor remaining at 0.185. A map calculated from the aquo-met model revealed positive difference peaks of 3 sigma height off the end of each of the oxygen atoms bound to the four heme irons, indicating that water alone cannot account for the observed electron density. These observations indicate that a diatomic ligand, presumably carbon monoxide, is predominantly bound to the Gower II hemoglobin.

## 4.4 SUBUNIT CONTACTS

The adult hemoglobin $\alpha_1\beta_1$ interface remains intact during the R to T transition of hemoglobin and is stabilized by a large number of interactions. In contrast the $\alpha_1\beta_2$ interface is where the quaternary switch occurs, with a different set of interactions formed in the T and R states. The $\alpha_1\varepsilon_1$ interface of Gower II hemoglobin contains three amino acid substitutions in the ε chain that alter intersubunit contacts with the corresponding α chain (Figure 4.11). The substitution at position 55, β Met → ε Leu results in the loss of a hydrophobic contact with the α119 Pro ring owing to the shorter side-chain of the leucine. Conversely the replacement of β112 Cys → ε112 Ile introduces a non-polar interaction between the εIle Cδ and α106 Cβ (with a distance between these atoms of 3.6 Å).

In the adult hemoglobin tetramer β116 His Nε2 forms a hydrogen bond with the carbonyl oxygen of α114 Pro (with the distance between these two atoms in the adult R and R2 structures being 2.9 Å and 2.7 Å respectively). The ε chain contains a threonine at this position with the side-chain Oγ hydrogen bonding back to the main-chain carbonyl oxygen of residue ε113, resulting in the loss of a hydrogen bond across the interface. The loss of a proline at position 125 (β125 Pro → ε125 Glu), on the fringe of the interface has little effect on the main-chain conformation, with the glutamate side-chain pointing into solvent.

The effects of these amino acid substitutions are unknown. The $\alpha_1\beta_1$ interface appears static during the quaternary switch, but it has been shown that oxidation of Met 55 in the adult β chain has a dramatic effect on the properties of the hemoglobin. The T state is destabilized resulting in a molecule showing a higher oxygen affinity and lower cooperativity (Amiconi et al., 1989).

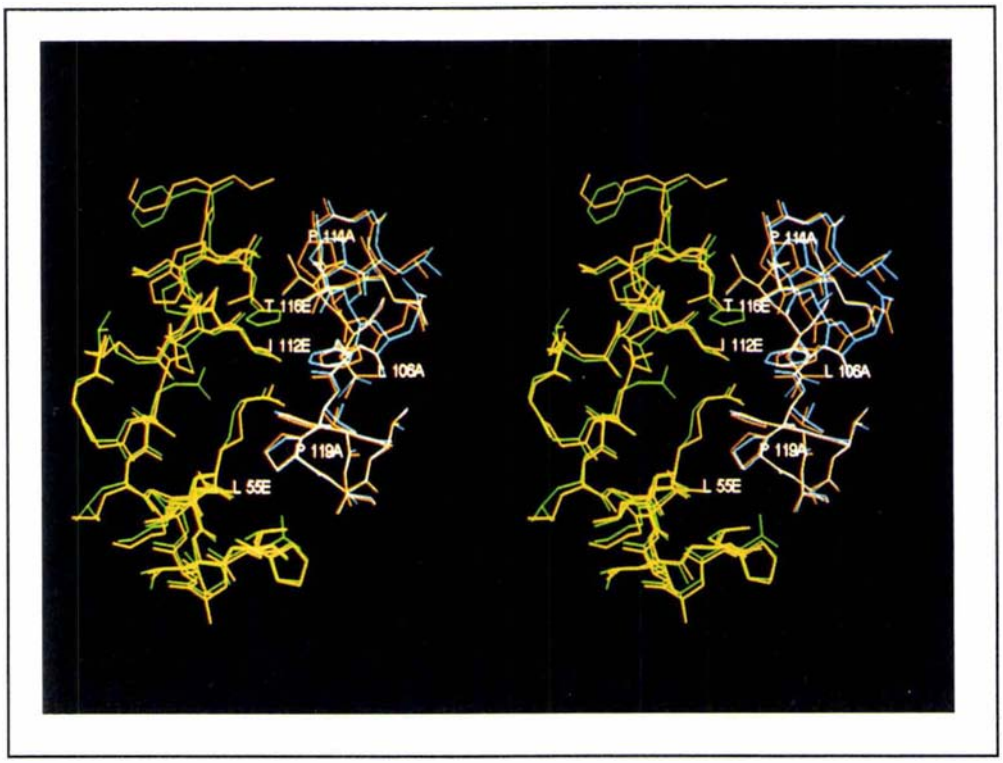



**Figure 4.11** Stereo superposition of the Gower II $\alpha_1\varepsilon_1$ and adult R2 $\alpha_1\beta_1$ interface

The Gower II Hb is coloured in orange ($\alpha$ chain) and yellow ($\varepsilon$ chain), with the adult Hb blue ($\alpha$ chain) and green ($\beta$ chain). The superposition was calculated on the BGH frame. The figure was prepared with TURBO-FRODO (Roussel & Cambillau, 1991).

The $\alpha_1\beta_2$ interface is more strictly conserved amongst members of the hemoglobin family. The only amino acid substitution occurring in this region is $\beta$43 Glu ➛ $\varepsilon$43 Asp (also aspartate in fetal $\gamma$ chain). In the epsilon monomer this aspartate side-chain forms an ionic interaction with the guanidinium group of $\alpha$92 Arg in a similar manner to the $\beta$ glutamate, resulting in little overall change. The degree of structural conservation in this region is reflected in the similar values obtained for the dimerization constants of the deoxy adult and Gower II hemoglobins (Hofmann and Brittain, 1996).

## 4.5 IMPLICATIONS FOR FUNCTION

The generally accepted physiological mode of action of the embryonic hemoglobins is that they strip oxygen from the maternal interstitial fluid to fulfil the oxygen demands of the respiring embryonic tissue. In order to achieve this the embryonic hemoglobin would be expected to have a higher affinity for oxygen than the adult protein *in vivo*. However, in the complete absence of any allosteric effectors the

three embryonic hemoglobins have a very similar oxygen affinity (Hofmann et al., 1995b) and binding rates relative to the adult protein (Hofmann and Brittain, 1996). Comparisons of the human adult, fetal and embryonic Gower II hemoglobin structures show that the proteins have very similar quaternary and tertiary conformations. However they differ enough to modify the oxygen binding properties sufficiently to allow the operation of these molecules in their physiological environment. It appears that the regions of the molecule essential for intrinsic hemoglobin function, i.e the heme pockets and the $\alpha_1\beta_2$ interface, are highly conserved, but it is the response to allosteric effectors in the prenatal hemoglobins that is reduced, resulting in a higher oxygen affinity under the same concentration of these effectors.

Chloride ions play an allosteric role in the regulation of hemoglobin affinity. Within the central cavity of the hemoglobin tetramer lies a destabilising excess of positive charge (Bonaventura and Bonaventura, 1978). Diffusion of chloride ions into the cavity results in the neutralization of this charge, though it appears that the chlorides do not bind to any specific sites (Perutz et al., 1993; Kavanaugh et al., 1992). Diffusion of chloride ions is greater in the T quaternary state owing to the larger width of the central cavity, resulting in a greater stabilization of this state and hence a lowering of the oxygen affinity. The polar residues contained within the central cavity of the adult hemoglobin tetramer are the anionic α94 Asp, α126 Asp and β101 Glu, and the cationic α1 Val amino-terminus, α99 Lys, α103 His, β1 Val amino-terminus, β2 His, β82 Lys, β104 Arg and β143 His. Despite the net surplus of five pairs of positively charged residues the adult tetramer behaves as though there is an excess of only two pairs of positive charged amino acids, implying that the histidine residues may not be completely ionized and that the lysine residues have reduced pKa values (Bonaventura et al., 1994; Perutz et al., 1994).

Oxygen binding experiments conducted in the presence of varying concentrations of chloride ions reveal that the Gower II hemoglobin displays a reduced sensitivity to the allosteric effects of chloride when compared to adult hemoglobin (Hofmann et al., 1995b), and it has hence been proposed that this is the principal effect governing the higher oxygen affinity of this molecule. The difference in chloride effect observed between the adult and Gower II hemoglobins is consistent with the removal of a pair of positive charges from the central cavity of the embryonic hemoglobin tetramer (Perutz et al., 1994), and hence one charge from the ε subunit, as the α chain is identical. The polar residues lining the central cavity are conserved

between the adult β and embryonic ε chains with the exception of 104 β Arg → ε Lys (Figure 4.12).

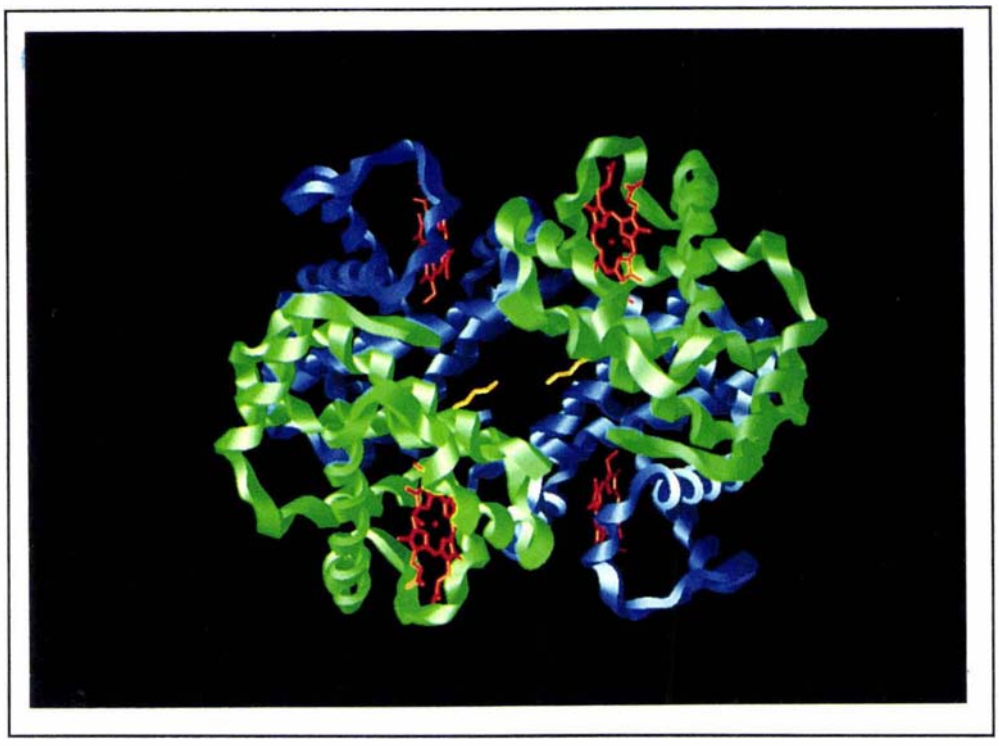


**Figure 4.12** Ribbon diagram of the Gower II Hb tetramer showing the position of ε104 Lys in the central cavity.

The α subunits are blue, and the ε subunits green. ε104 Lys is shown as a yellow stick representation. This figure was prepared with GRASP (Nicholls et al., 1991).

In the Gower II structure ε104 Lys is involved in an ionic interaction with ε101 Glu (Figure 4.13), though density is weak for the lysine side-chain which may be partially disordered; this is common for lysine and arginine side-chains contained within the central cavity. In the adult R and R2 structures β104 Arg forms hydrogen bonds with the main-chain carbonyl oxygens of β100 Pro and β139 Asn as well as the side-chain carbonyl of β139 Asn (Ile in ε). In the T quaternary structure β104 Arg interacts with β101 Glu. The 104 β Arg → ε Lys substitution results in the replacement of a residue with one having an intrinsically lower pKa. If the environment of the ε104 Lys was such that the pKa was lowered enough so that at physiological pH the amine was unionized, then this substitution could explain the reduction in the chloride effect observed for Gower II hemoglobin. However, examination of the environment around this residue provides no obvious insights as to whether any reduction in pKa is likely. Bovine hemoglobin also has a lysine residue at the equivalent position, with this hemoglobin showing a response to chloride ions similar to that of human adult

hemoglobin. It is much more difficult, however, to draw conclusions from these two proteins as bovine hemoglobin shows a three-fold decrease in its intrinsic oxygen affinity and also has β2 His deleted (Perutz et al., 1993).

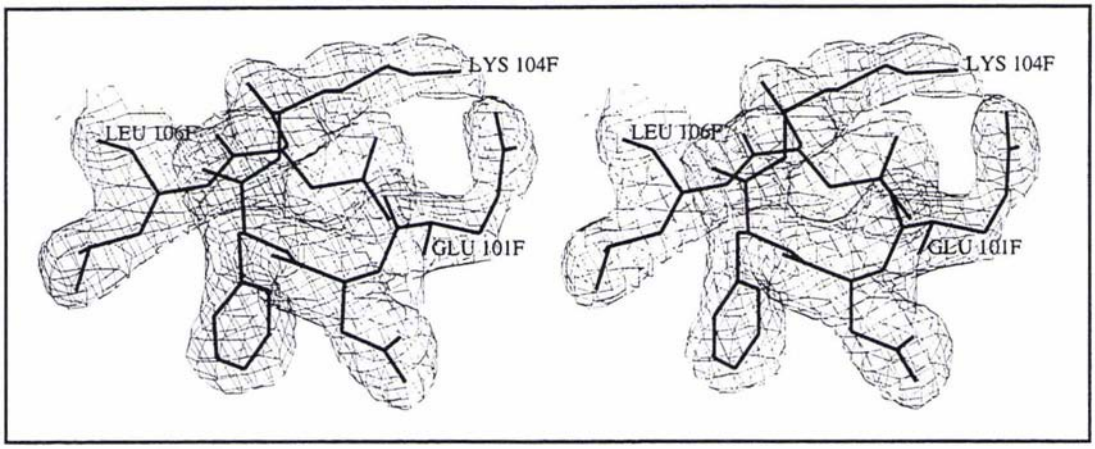



**Figure 4.13** The environment of Lys 104 in the ε chain.

The $2m|F_o| - D|F_c|$ map is contoured at 1 sigma. This figure was prepared with TURBO-FRODO (Roussel & Cambillau, 1991).

Three other amino acid substitutions between the ε and β had been suggested as possible changes responsible for reducing the net positive charge inside the cavity and altering the chloride effect: β77 His → ε77 Asn, β116 His → ε116 Thr, and β125 Pro → ε125 Glu (Hofmann et al., 1995b). Though the Gower II structure is liganded, it is possible to infer the approximate position of these residues in the T state by comparison with adult T state models. ε116 Thr is situated at the $\alpha_1\beta_1$ interface and is buried in the area forming the dimer. This region of the structure is invariant during the R to T transition and is unlikely to be accessible to chloride ions. ε77 Asn is located at the C-terminus of the E helix, with the asparagine side-chain on the surface of the molecule, while ε125 Glu is positioned at the N-terminus of Helix H, with its side-chain also pointing out into solvent. Neither of these two residues are located within the central cavity, so any overall charge differences contributed are located on the surface of the molecule. It is unclear as to whether the reduced Gower II chloride effect can be simply explained by a single amino acid substitution. Rather it may be a combination of smaller electrostatic changes that cause the differences observed.

2,3-bisphosphoglycerate lowers the oxygen affinity by binding at the top of the central cavity of the T state hemoglobin tetramer between the N and C-termini of the two

β chains, thus stabilizing this quaternary form. The residues which form the 2,3BPG binding site, i.e. 1-α-amino terminus, 2His, 82Lys and 143His, are conserved between the β and ε chains. Oxygen binding experiments show that Gower II hemoglobin exhibits a reduced allosteric response to 2,3BPG, in comparison to the adult molecule (Hofmann et al., 1995a). It cannot be predicted whether the shift observed in the A helix and N-terminal region of the ε chain in the Gower II carbonmonoxy structure is preserved after the quaternary switch and hence will have any influence on 2,3BPG binding. This N-terminal shift is also observed in the γ subunit of fetal deoxyhemoglobin and has been proposed to alter the affinity of fetal hemoglobin for 2,3BPG, in conjunction with the143β His ↣ 143γ Ser substitution which removes two of the phosphate binding ligands (Frier and Perutz, 1977). The altered binding of 2,3BPG to fetal hemoglobin is thought to be the predominant cause of the higher oxygen affinity *in vivo* (Bauer et al., 1968; Tyuma and Shimizu, 1970).

During prenatal human development a system of oxygen supply has evolved enabling the embryo to obtain oxygen from the mother. The Gower II hemoglobin molecule has an extremely similar conformation to that of the adult hemoglobin, but subtle differences exist that allow the function of this molecule *in vivo*. There exists a reduced response to chloride ions of the embryonic hemoglobins, the extent of which depends on the subunit composition of the particular molecule, and an altered effect of 2,3BPG binding on the fetal and Portland embryonic hemoglobin which both contain the γ subunit. However, the comparative affinities of the hemoglobins *in vivo* are also dependent on the relative amounts of these allosteric effectors in the maternal and embryonic red blood cells.

## 4.6 SUGGESTIONS FOR FUTURE STUDIES

A structural characterization of the Gower II human embryonic hemoglobin in the T state is a priority. It has been proposed that the destabilization of this quaternary form is responsible for the higher oxygen affinity under presumed physiological conditions (Brittain et al., 1997). The structural investigation of 2,3BPG binding could be conducted, and comparisons made with the interactions of the adult molecule with this allosteric effector. A structure of the T state Gower II hemoglobin is also essential for a full investigation of the origins of the decreased sensitivity of this protein to chloride ions, with this effect hypothesized as being the predominant cause of a higher oxygen affinity of this molecule *in vivo.* The structure of the liganded Gower II hemoglobin, as well as other structural studies (Frier and Perutz, 1977), has shown that the overall folds of the human hemoglobins are very similar. In order to

fully characterize any structural differences that may account for the observed alterations in functional behaviour high resolution studies, if possible, would be most appropriate.

Investigation of the chloride effect could be enhanced by site directed mutagenesis of the Gower II hemoglobin in conjunction with both functional and structural studies. It would be interesting to determine whether any change in chloride affinity is observed by mutating ε104 Lys. Suitable substitutions would include Lys → Arg (as found in the β chain), as well as substitutions to both neutral and negatively charged residues. The contributions of the amino acid substitutions between the ε and β chains suggested to be responsible for the altered chloride effect (ε77 Asn, ε116 Thr, and ε125 Glu, Hofmann et al., 1995a) could also be investigated by mutagenic studies. It might also be possible to modify ε104 Lys by chemical means (e.g. by the action of pyridoxal 5'-phosphate) and hence remove the contribution (if any) of this side-chain to the cationic excess of the tetramer central cavity.

In order to full understand oxygen transport in the developing embryo it is imperative that structural information be obtained for the Gower I and Portland embryonic hemoglobins. There is considerable interest in the structure of the embryonic ζ chain, argued to be a 'primitive' chain more highly diverged from the other members of the human hemoglobin family (Melderis et al., 1974). The level of sequence similarity for this chain is relatively low when compared to other members of the generally well conserved globin family. The ζ chain has more sequence similarity to the embryonic globin chains from chickens, suggesting the divergence of this gene from the adult α chain prior to the divergence of birds and mammals (Proudfoot et al., 1982). The structure of the N-terminus of ζ chain is of importance as it is the only human globin chain that contains a modified N-terminus.

## 4.7 APPENDIX

### *4.7.1 Crystallization experiments on the Gower I and Portland embryonic hemoglobins*

Crystallization experiments have been conducted on the Gower I ($\zeta_2\varepsilon_2$) and Portland ($\zeta_2\gamma_2$) hemoglobins. A broad screen of crystallization conditions for the carbonmonoxy Portland hemoglobin has been undertaken utilizing screens based empirically on previously successful crystallization conditions (Jancarik and Kim, 1991) and those of a factorial design (Kingston et al., 1994). Crystallization experiments were also performed under the high phosphate conditions from which adult hemoglobin is crystallized (Perutz, 1968). The Portland hemoglobin exhibited a decreased solubility compared to the Gower II protein. The only crystals that appeared from these screening experiments were grown from 1.0M sodium acetate at pH 6.5 (buffered by 0.1M imidazole) at room temperature. The bright red crystals were very small and had an irregular appearance, with no crystalline faces apparent (Figure 4.14). The crystals appeared over a period of around a week. A systematic search around these initial conditions revealed that these crystals could be grown in the pH range 6.5 to 8.0, though it was essential that imidazole was the buffer used. The sodium acetate concentration could be varied between 0.4 and 1.4M with a protein concentration between 2 and 10 mg/ml. No improvement in crystal quality or size resulted from these experiments. Crystallization in the presence of glycerol was investigated. Concentrations of between 5 and 20%(v/v) appeared to have little observable effect on crystal growth, though higher concentrations (>30%) resulted in an increased number of even smaller crystals. These crystallization experiments were conducted using hanging drops. On appearance of the crystals, a thick skin had developed over the surface of the drop. The skin was present regardless of whether glycerol had been included in the crystallization mother-liquor. The formation of a skin over a hanging drop is not uncommon for crystallization experiments involving polyethylene glycol polymers, but was unexpected for a system comprising salt and buffer. Imidazole is known to oxidize hemoglobin to the met form, and excess imidazole can result in denaturation of the protein (Di Iorio, 1981). Hence it may be the case that the crystals are a heme derivative and that the thick skin enveloping the hanging drops is denatured globin. It has not been possible to separate the crystals from the skin for further analysis. All further attempts to obtain crystals suitable for X-ray diffraction experiments have proved to be unsuccessful.

Initial crystallization screening experiments have also been conducted on the carbonmonoxy Gower I hemoglobin. These trials have been unsuccessful at identifying conditions suitable for crystal growth. Crystallization experiments on both these embryonic hemoglobins are continuing.

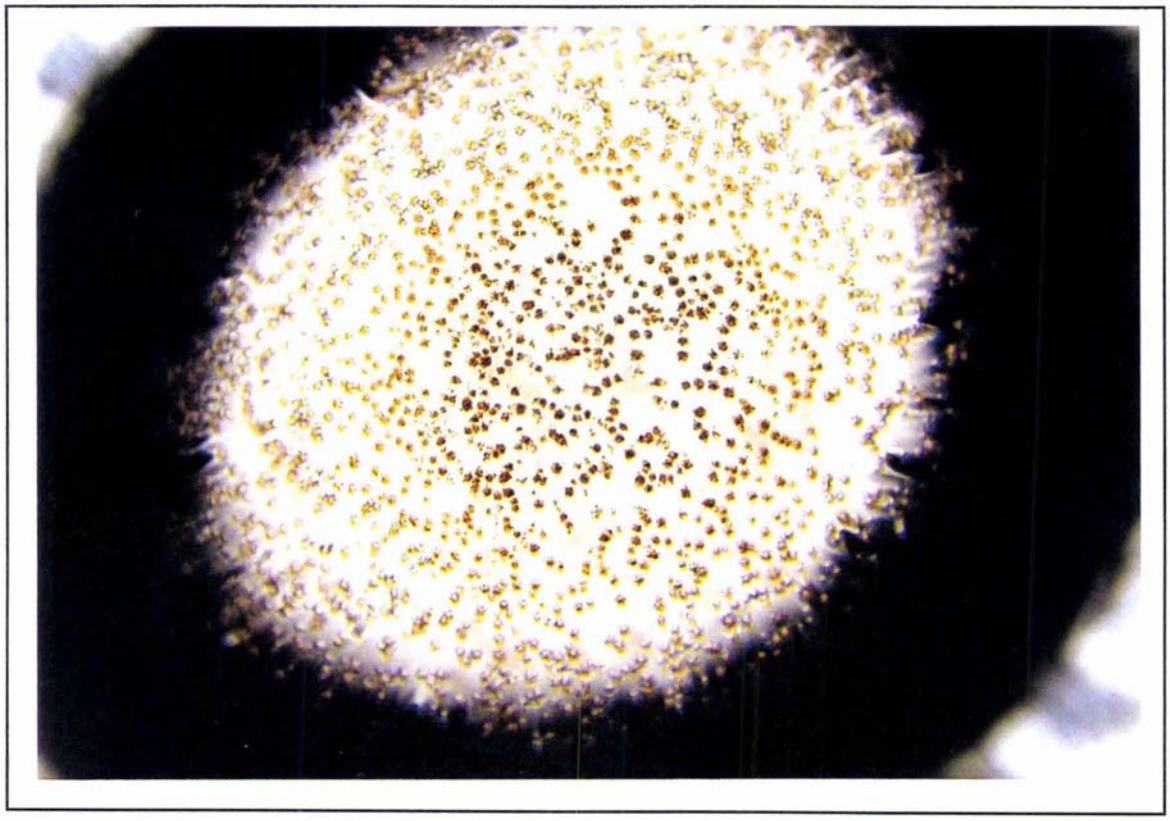


**Figure 4.14** Portland hemoglobin crystals

The total magnification is ~100x

Chapter 5

# *P. stutzeri* Nitrite Reductase: An Overview

## 5.1 THE PSEUDOMONAS GENUS

Bacteria of the *Pseudomonas* genus have straight or curved morphology and a size of between 0.5-1.0μm by 1.5-4.0μm. Some members of this genus can utilise nitrate as an anaerobic electron acceptor. These bacteria grow chemoorganotrophically at neutral pH at mesophilic temperatures, usually in soil and sediments. Some species of pseudomonads are human pathogens, e.g. *P. aeruginosa* is often found in hospitals associated with infection of human urinary and respiratory tracts and can opportunistically invade patients with severe burns and other skin damage. *P. stutzeri* is also often isolated from humans, living saphrolytically in the body. *Pseudomonas* species can also act as plant pathogens. e.g. *P. marginalis, P. syringae*, and *P. solanacearun.* Recent interest in this genus has been stimulated by the discovery of a 'superbug' strain of *P. aeruginosa* that is resistant to the carbapenems, a powerful class of antibiotics and the current drug of last resort (Day, 1998).

Some pseudomonads have the ability to behave as facultative anaerobes, utilising an alternative electron acceptor in the absence of oxygen. Anaerobic respiration is less efficient, but enables the bacteria to inhabit a wider range of environmental conditions.

Bacterial denitrification is the pathway in which nitrogen oxide species are utilised as electron acceptors in the place of oxygen. Electron transport phosphorylation is coupled to the enzymatic reduction of the nitrogen oxides $NO_3^-$, $NO_2^-$, NO, and $N_2O$ to $N_2$. This process is the major means by which inorganic nitrogen is converted to nitrogen gas and returned to the atmosphere (Figure 5.1).

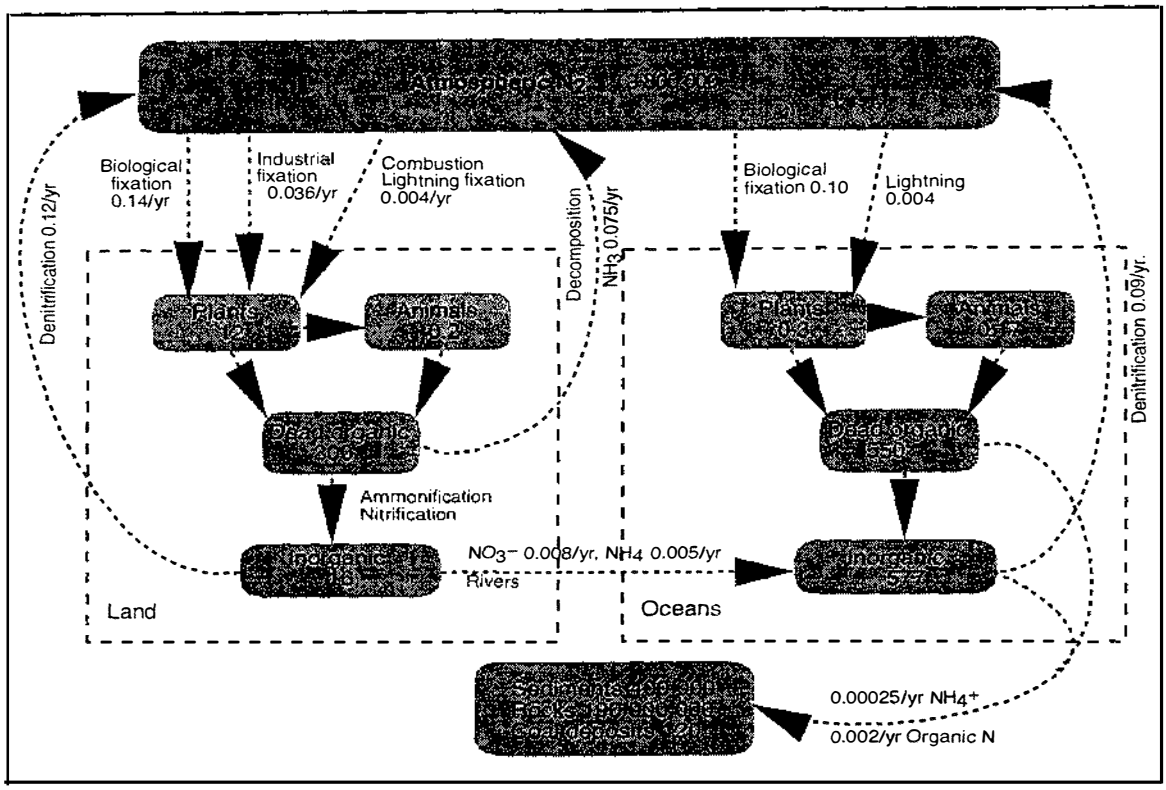



**Figure 5.1** Schematic diagram of the global nitrogen cycle.

Numbers given are in units $x10^{15}$ grams. Modified from Brock et al., 1997

It is for this reason that the pseudomonads are considered important microbes in soil and water. Bacterial denitrification has important consequences for agriculture as it causes the removal of applied nitrogenous fertilizer from the soil resulting in decreased crop productivity. It has been estimated that denitrification results in the loss of as much as 20 to 30% of introduced nitrogen from the soil (Firestone, 1982). Denitrification is not a completely sealed system in all species resulting in the release of nitric and nitrous oxides, both of which are greenhouse gases implicated in global warming and destruction of the ozone layer. Bacterial denitrification has found practical applications in biotechnology and environmental sciences. The presence of denitrifying bacteria in sewerage and wastewater is beneficial, decreasing the levels of available nitrogen that stimulates algal growth. Several environmental pollutants such as phenol (Tschech and Fuchs, 1987), toluene, 1,3-dimethylbenzene (Kuhn et al., 1988), carbon tetrachloride (Criddle et al., 1990) and nitrilotriacetic acid (Wanner et al., 1990) have been reported to be broken down under denitrifying conditions. Intrinsic interest has been high in the enzymes that catalyze

the reductive steps. The nitrogen oxide reductases utilise a wide range of metallic cofactors: molybdopterin, mononulear and dinuclear copper centres, heme groups, iron-sulfur clusters and possibly non-heme iron.

Certain bacteria e.g. *Wollinella succinogenes* and *Vibrio fischeri* can also utilise reduced nitrate as a nitrogen source for growth. Nitrate is reduced to ammonia in a process called assimilative nitrate reduction and secreted into the medium, during which no nett nitrogen is lost from the immediate microbial environment. This reaction is catalyzed by a 50-60 kDa hexa-heme protein (Liu et al., 1988). Plants and certain other bacteria carry out a similar reduction in which the ammonia produced is utilised for biosynthetic purposes and not secreted. These nitrate reductases contain an iron-sulphur cluster and a siroheme group (Brittain et al., 1992). These pathways are in contrast to dissimilative nitrate reduction or denitrification.

## 5.2 OVERVIEW OF BACTERIAL DENITRIFICATION

Denitrification was first identified as a microbial process in 1882 (Gayon and Dupetit, 1882). Since this time the denitrification pathway has been studied in predominantly two bacterial species: the membrane-bound nitrate reductase and cytochromes have been investigated in *Paracoccus denitrificans* (Berks et al., 1995a), whereas the soluble nitrite reductase from *Pseudomonas aeruginosa* has been the subject of much investigation (reviewed in Henry and Bessieres, 1984; Silvestrini et al., 1994), with more recent research on the enzymes from *Thiosphaera pantotropha* and *Pseudomonas stutzeri.*

The sequence of reductive steps in bacterial denitrification is:

$$NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2$$

The four enzymes responsible for catalyzing each redox step are rather predictably named: nitrate reductase, nitrite reductase, nitric oxide reductase and nitrous oxide reductase. Some bacteria appear to lack a nitrous oxide reductase and hence the pathway stops with $N_2O$ as the final product. Both the nitrite and nitrous oxide reductases are soluble periplasmic enzymes, with the exception being the membrane associated nitrite reductase isolated from *T. denitrificans* (Hole et al., 1996), whereas the nitrate and nitric oxide reductases are integral membrane proteins. Several reviews have been written in recent years on the subject of bacterial deni-

trification (Ye et al., 1994; Berks et al., 1995a; Averill, 1996). The nature of bacterial periplasm is reviewed by Ferguson (Ferguson, 1992).

Bacteria employ the standard tricarboxylic acid cycle, with the electrons donated to the electron transport chain coming from NADH and succinate. The membrane-bound nitrate reductase receives electrons from the ubiquinol pool via a redox process that is coupled to the generation of a transmembrane proton gradient (see Figure 5.2). The reduction of nitrate to nitrite occurs at the $\alpha$ subunit on the cytoplasmic side of the membrane. A consequence of this is that nitrate obtained from outside the cell has to be transported across the membrane against the membrane potential, and subsequently nitrite delivered to the periplasm, the location of the nitrite reductase. The mechanism of these delivery processes is not understood, but it is known that the transport mechanism is inhibited by oxygen and can discriminate against chlorate as substrate (Alefounder and Ferguson, 1980). The nitrite is reduced by the periplasmic nitrite reductase utilising electrons delivered from the membrane-bound cytochrome $bc_1$ via a periplasmic cytochrome *c* or blue copper protein (azurin or pseudoazurin). The NO produced by this enzyme undergoes further reduction by the membrane-bound nitric oxide reductase to give $N_2O$. In certain bacteria this is the final product of denitrification, but generally the $N_2O$ undergoes further reduction to $N_2$ by the periplasmic nitrous oxide reductase. The electrons for these two reduction steps are also delivered by the same mechanism as that for nitrite reductase.

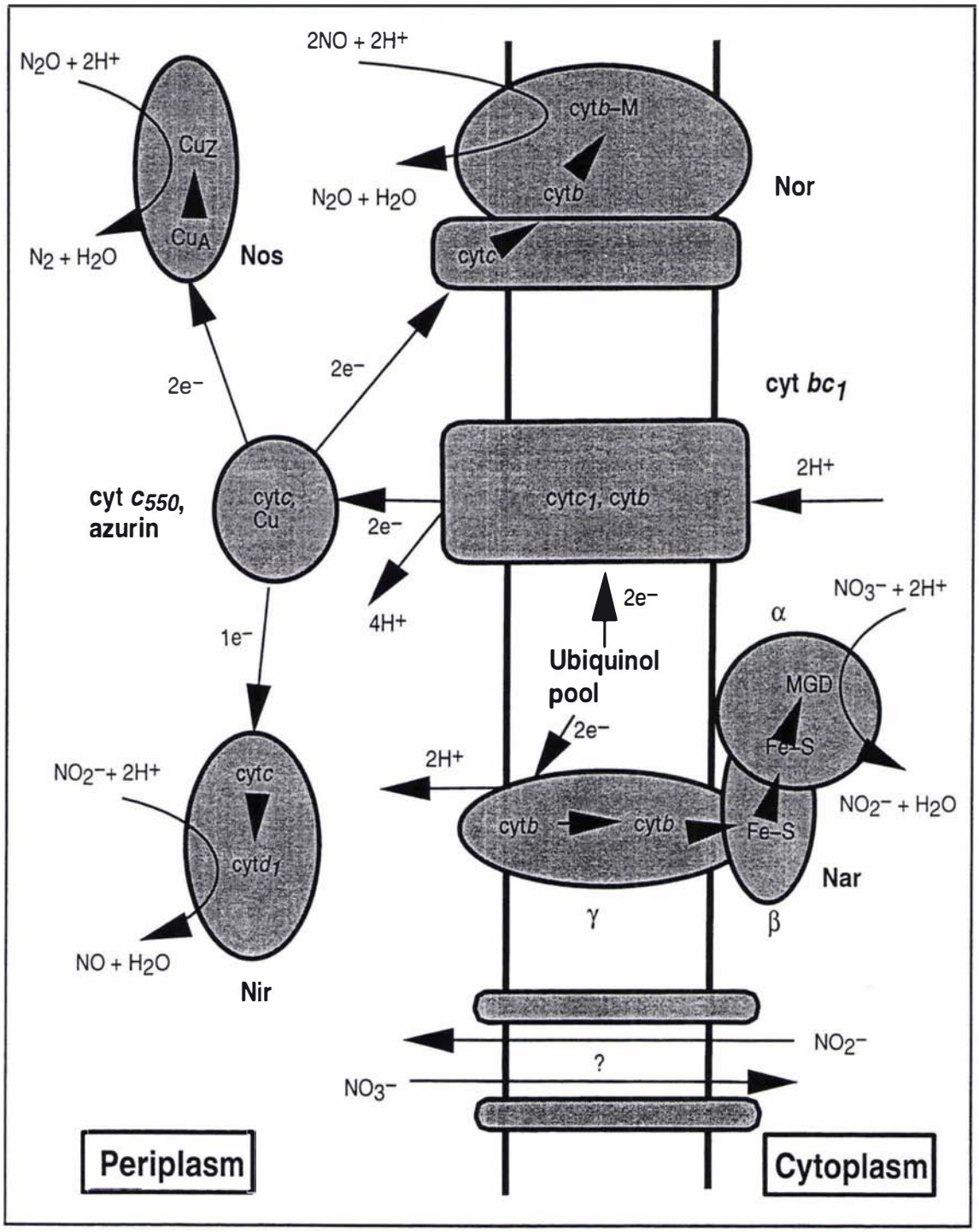



**Figure 5.2** Schematic overview of bacterial denitrification

Nir=nitrite reductase, Nar=nitrate reductase, Nor=nitric oxide reductase, Nos= nitrous oxide reductase, cyt=cytochrome. Modified from Berks et al., 1995a

## 5.3 CYTOCHROME $cd_1$-NITRITE REDUCTASE

### *5.3.1 Introduction*

A soluble cytochrome oxidase was first isolated and characterised from *Pseudomonas aeruginosa* grown anaerobically in the presence of nitrate by Horio and coworkers (Horio et al., 1960). They also isolated the two redox proteins, cytochrome $c_{551}$ and azurin that were able to act as electron donors to the oxidase. The enzyme was initially characterised for its oxygen reductase activity, reducing oxygen to water.

$$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O$$

Because of this activity and the fact that the enzyme was inhibited by CO and $CN^-$ this soluble bacterial cytochrome oxidase was considered to be a suitable model for the membrane bound cytochrome oxidase. A short time later Walker and Nicholas (Walker and Nicholas, 1960) reported the discovery of a yellow-green protein that catalysed the reduction of nitrite to nitric oxide.

$$NO_2^- + 2H^+ + e^- \rightarrow NO + H_2O$$

Yamanaka and coworkers made similar observations (Yamanaka et al., 1961) and also reported that anaerobic conditions were not the essential factor in nitrite reductase activity, but the presence of $KNO_3$ in the growth media. Anaerobic conditions, though not being essential, do favour nitrite reductase activity (Parr et al., 1976). Kinetic and equilibrium measurements have lead to the conclusion that nitrite reduction is the physiologically relevant reaction of this enzyme (Silvestrini et al., 1990). Hence this enzyme, formerly called cytochrome oxidase, is now referred to as cytochrome $cd_1$ nitrite reductase. This class of enzyme is distinct from the Cu containing nitrite reductases (Section 5.4.2).

Approximately two-thirds of denitrifying bacteria examined to date contain a cytochrome $cd_1$ nitrite reductase, with the remainder having a copper containing enzyme to catalyse this reaction. Cytochrome $cd_1$ enzymes have been isolated from a range of bacteria. including *Pseudomonas stutzeri*, *Pseudomonas aeruginosa*, *Thiobacillus denitrificans*, *Alcaligenes faecalis*, *Pseudomonas halodenitrificans*, *Paracoccus denitrificans* and *Thiosphaera pantotropha* (Averill, 1996).

### 5.3.2 Heme groups and their biological role

The cytochrome $cd_1$, as its name implies, contains one *c* heme and one $d_1$ heme per monomer. The *c* heme, as is found in all *c*-type cytochromes, is covalently bound to the polypeptide chain via two thioether linkages to cysteine residues. These are contained in a conserved structural motif, Cys-X-X-Cys-His, with the histidine side-chain providing one of the axial heme ligands. The class I cytochrome *c* type of coordination, hexacoordinated and low spin with histidine and methionine axial ligands, had been proposed on the basis of magnetic circular dichroism and EPR spectroscopic studies on 'resting' oxidised nitrite reductase from *P. aeruginosa* (Walsh et al., 1979; Sutherland et al., 1986). More recent studies on the enzyme from *P. stutzeri* have yielded similar results (Cheesman et al., 1997), in contrast to measurements on the oxidised enzyme from *T. pantotropha* which indicated that the *c* heme was coordinated by two histidine residues. The latter result was in agreement with the unexpected bis-histidine ligation of the *c* heme found in the crystal structure of the oxidised *T. pantotropha* enzyme (Fulop et al., 1995).

The $d_1$ heme is unique to the cytochrome $cd_1$ nitrite reductase enzymes and is the chromophore that is responsible for the distinctive olive green colour of the enzyme. Spectroscopic studies in conjunction with synthesis of model compounds led Chang and coworkers to propose a porphyrindione (dioxoisobacteriochlorin) structure for $d_1$ heme (Chang et al., 1986). Later crystal and NMR structures of heme $d_1$ (Barkigia et al., 1992) and more recently the crystal structures of nitrite reductase from *T. pantotropha* (Fulop et al., 1995) and *P. aeruginosa* (Nurizzo et al., 1997) have confirmed these conclusions. The $d_1$ heme is more bulky than *b* heme (Figure 1.4). Spectroscopic studies have shown that the ferric $d_1$ hemes are hexacoordinated, with those from *P. aeruginosa* and *P. stutzeri* low spin species, whereas the *T. pantotropha* cofactor is a low spin/high spin mixture. Thus it appears that the enzyme from *T. pantotropha* is significantly different to the enzymes studied from the *Pseudomonas* genus (Cheesman et al., 1997).

The reduced $d_1$ heme is a high spin pentacoordinated species in all cytochrome $cd_1$ enzymes studied, being able to bind ligands such as $NO_2^-$, NO, CO and $O_2$. The $d_1$ heme can be extracted from the enzyme via treatment with acidic acetone to give a semi-apoenzyme whose activity can be restored by reconstitution with synthetic $d_1$ heme (Hill and Wharton, 1978; Weeg-Aerssens et al., 1991). Reconstitution with

heme *a* results in an enzyme exhibiting 5% oxidase activity of the native. Under reducing conditions or when complexed with NO or CO there is a spontaneous, but slow, loss of the $d_1$ heme from the enzyme (Silvestrini et al., 1994).

The *c* heme's role is to accept electrons from the external donor proteins, whereas the $d_1$ heme is the site of nitrite or oxygen reduction. In *P. aeruginosa* it has been shown that cytochrome $c_{551}$ and azurin can act as electron donors (Parr et al., 1977; Silvestrini et al., 1982). A major question in understanding the function of this enzyme is what properties the unique $d_1$ heme provides the nitrite reductase, as opposed to a more ubiquitous heme cofactor like *b* heme.

### *5.3.3 Gene structure and regulation*

In *Pseudomonas stutzeri*, *Pseudomonas aeruginosa*, and *Paracoccus denitrificans* the gene encoding cytochrome $cd_1$ nitrite reductase (*nirS*) is clustered with genes encoding enzymes that may play a role in heme $d_1$ synthesis, as well as genes for other electron transport proteins (Figure 5.3).

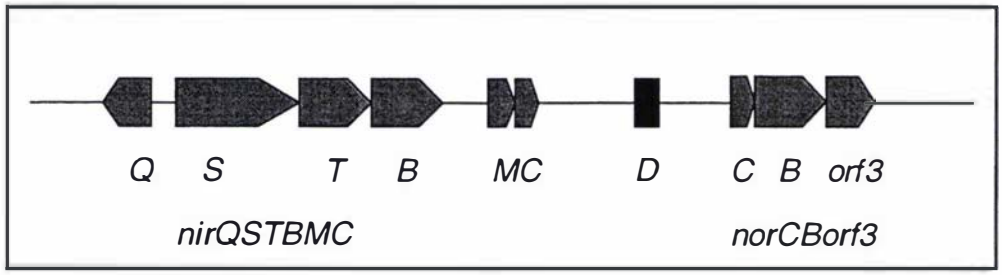



**Figure 5.3** Schematic diagram of the *P. stutzeri* gene cluster encoding cytochrome $cd_1$ and nitric oxide reductase.

Derived from Berks et al., 1995a.

In *P. stutzeri, nirS* is followed by genes for *nirT* - a tetra-*c* heme electron transport protein, *nirB* - a diheme cytochrome $c_{552}$, *nirM* - a monoheme cytochrome $c_{551}$, and *nirC* - a small monoheme *c*-type cytochrome (Jungst et al., 1991b). The nitrite reductase loci from *P. aeruginosa* (Silvestrini et al., 1989) and *P. denitrificans* (de Boer et al., 1994) do not contain the *nirT and nirB* genes. The *nirD* gene located further downstream has not been analysed in detail, but appears to be necessary for

cytochrome $cd_1$ synthesis. Mutations in this gene result in synthesis of an inactive nitrite reductase lacking $d_1$ heme (de Boer et al., 1994). The genes encoding the two subunits of nitric oxide reductase (*norC* and *norB*), and also an unknown cytoplasmic protein (*orf3*), are located near the nitrite reductase gene cluster (Jungst et al., 1991a).

The regulation mechanism of these genes is not well understood, though it is clear that both oxygen and N-oxides can control expression. There appears to be some variability amongst the different denitrifying bacteria studied. In general these genes are not expressed under aerobic conditions, resulting in the energetically more favourable respiration of oxygen occurring in preference to nitrogen oxide respiration. There is also evidence that the genes encoding all the enzymes for denitrification are coregulated to a certain degree (de Boer et al., 1994). Arai and co-workers report that in *P. aeruginosa* the expression of the *denAB* (nitrite reductase) gene is regulated in two steps, induction by nitrite and repression by oxygen (Arai et al., 1991a; Arai et al., 1991b). Korner and Zumft have shown that anaerobiosis by itself does not induce gene expression in *P. stutzeri*, but the presence of an N-oxide is essential along with anaerobic conditions (Korner and Zumft, 1989). There is circumstantial evidence of an *fnr* like transcription factor in *P. stutzeri* that may play a role in gene regulation (Cuypers and Zumft, 1993). Robertson and colleagues have shown that denitrification and oxygen respiration can occur at the same time in their studies of *T. pantotropha* (Robertson and Kuenen, 1984).

### *5.3.4 Protein sequence*

To date there have been six sequences reported of cytochrome $cd_1$ nitrite reductases. These sequences are from two strains of *P. stutzeri* (JM300 (Smith and Tiedje, 1992) and ZoBell (Jungst et al., 1991a)), *P. denitrificans* (de Boer et al., 1994) and its sub-species *T. pantotropha* (Baker et al., 1997), as well as *P. aeruginosa* (Silvestrini et al., 1989), and *Alcaligenes eutrophus* (Rees et al., 1997). The sequences exhibit a high degree of variability in their N-terminal regions. All the polypeptides contain the *c* heme-binding motif common to the class I cytochromes, Cys-X-X-Cys-His. The degree of sequence similarity is greater in the C-terminal part of the enzyme. Examined in a pairwise manner the *P. stutzeri* JM300 and ZoBell nitrite reductases exhibit 91% sequence identity (93% for the N-terminal domain and 91% for the C-terminal domain). When the *P. stutzeri* JM300 nitrite reductase is compared to the *P. aeruginosa* and *T. pantotropha* enzymes the values are 54% (50% for the N-terminal domain and 55% for the C-terminal domain) and

52% (40% for the N-terminal domain and 55% for the C-terminal domain) respectively. It is not possible to identify the $d_1$ heme ligands by examination of the sequences alone. Consistent with its location in the periplasm the *P. stutzeri* nitrite reductase contains a 26-residue signal sequence for export of the protein from the cytoplasm (Jungst et al., 1991a). All the nitrite reductase sequences determined to date indicate the absence of disulphide bonds for these enzymes. The only cysteines contained within the enzyme are involved in the covalent attachment of the *c* heme. This is in keeping with a protein that is exported to the periplasm under anaerobic conditions and hence is not exposed to an oxidizing environment (Ferguson, 1992).

In all future descriptions, *P. stutzeri* nitrite reductase will refer to the enzyme from strain JM300 unless otherwise specified.

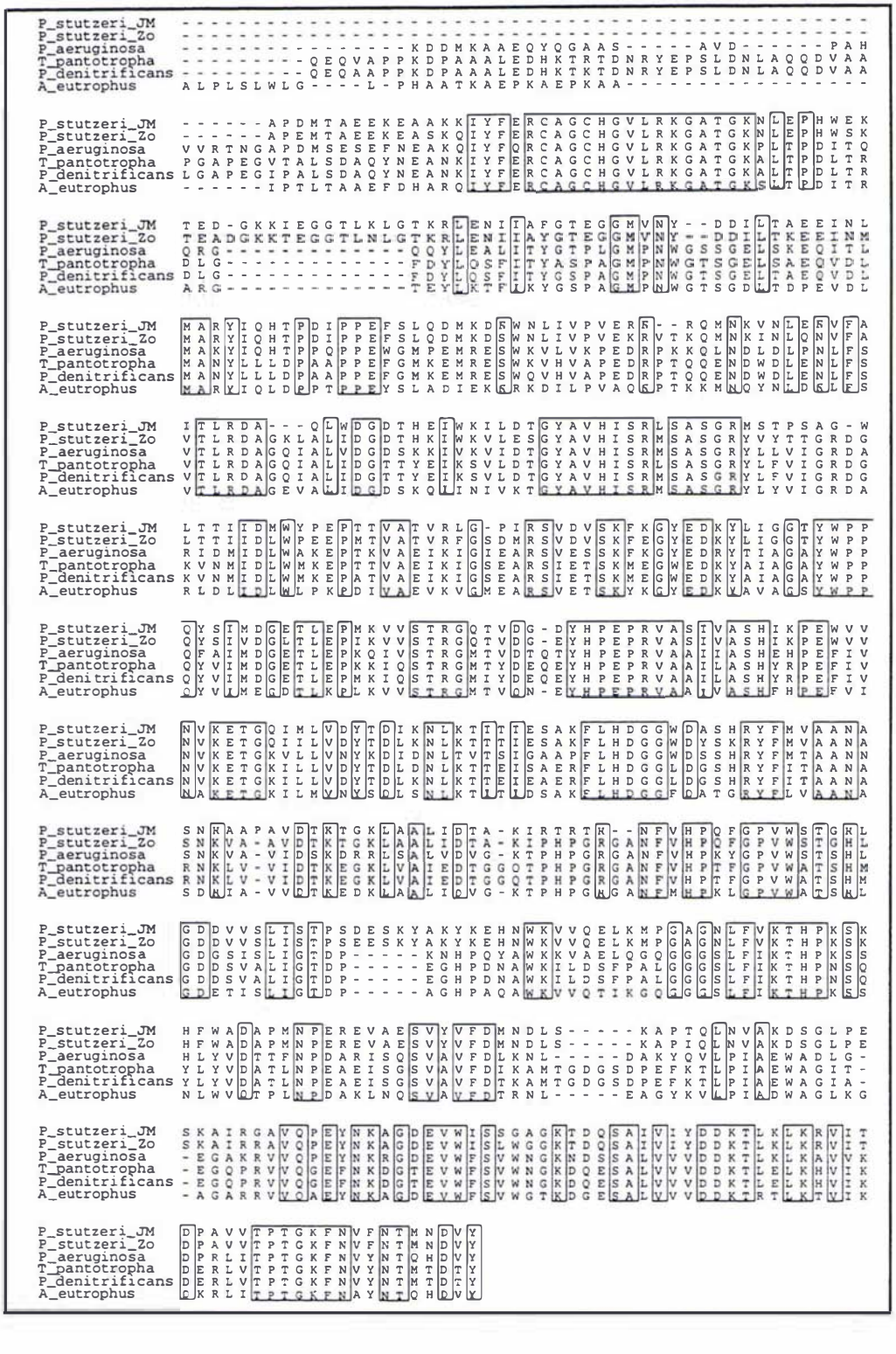



**Figure 5.4** Sequence alignment of the six known cytochrome $cd_1$-nitrite reductase protein sequences

JM and Zo are abbreviations for JM300 and ZoBell respectively. Identical residues in the sequences are boxed. Alignment carried out with CLUSTALW (Thompson et al., 1994) and displayed using ALSCRIPT (Barton, 1993).

### 5.3.5 Molecular structure

The cytochrome $cd_1$ nitrite reductases are homodimeric enzymes with a subunit molecular mass of around 60 kDa, as determined by sedimentation and chromatographic analysis (Barber et al., 1976). An approximate model of the oxidised dimer as an ellipsoid of dimensions 40 Å by 20 Å has been elucidated by both electron microscopy (Saraste et al., 1977) and small-angle X-ray scattering (Berger and Wharton, 1980). When the enzyme is reduced, however, these dimensions change substantially to 30 Å by 25 Å.

Recently three crystallographic analyses have been conducted on cytochrome $cd_1$ enzymes, the reduced (Williams et al., 1997) and oxidised (Fulop et al., 1995) states of the enzyme from *T. pantotropha*, and the oxidised *P. aeruginosa* nitrite reductase (Nurizzo et al., 1997). The overall architecture of the enzymes is similar, though substantial differences occur, particularly in the nature of the heme ligands between the enzymes from different species. Discussion and comparison of these structures and the *P. stutzeri* JM300 nitrite reductase model are provided in Chapter 7.

Briefly, the protein structure is divided into two domains. The smaller N-terminal domain, containing the *c* heme, resembles the type I cytochrome *c* proteins e.g. cytochrome $c_{551}$ from *P. aeruginosa* (Matsuura et al., 1982; Timkovich and Cai, 1993), and tuna cytochrome *c* (Takano and Dickerson, 1981). The larger C-terminal domain is an eight-bladed β-propeller structure. The $d_1$ heme is nested on top, in the centre of the β-propeller domain.

The oxidised *T. pantotropha* structure has two histidine axial ligands to the *c* heme, whereas the reduced enzyme and the oxidised *P. aeruginosa* nitrite reductase have the histidine/methionine coordination typical of the type I *c* cytochromes. Implications for the reaction mechanism will be discussed later. The $d_1$ heme in the oxidised *T. pantotropha* structure also exhibits unusual ligation having one histidine axial ligand and one tyrosine ligand which is provided by the N-terminal *c* heme domain. The tyrosine ligand is replaced, in the oxidized *P. aeruginosa* nitrite reductase, by an $OH^-/H_2O$ ligand which is hydrogen bonded to a homologous tyrosine.

### 5.3.6 Nitrite reduction reaction mechanism

The oxidised nitrite reductase is regarded as the resting state of the enzyme. For catalysis the *c* heme is reduced first, followed by an internal electron transfer to reduce the $d_1$ heme, allowing nitrite or oxygen to bind.

The transfer of electrons between the macromolecular electron donors azurin and cytochrome $c_{551}$ has been investigated both under transient and steady state conditions. In *P. aeruginosa*, cytochrome $c_{551}$ has a higher affinity than azurin for nitrite reductase. However both reduction rates are faster than the rate determining step for nitrite reduction, suggesting either electron donor may function *in vivo* (Tordi et al., 1985). It has been proposed that azurin binds via a hydrophobic patch (Van de Kamp et al., 1990) and cytochrome $c_{551}$ via an exposed heme edge (Silvestrini et al., 1982). The docking of pseudoazurin and cytochrome $c_{550}$ to cytochrome $cd_1$ from *T. pantotropha* has been modelled (Williams et al., 1995). A hydrophobic patch and surrounding positively charged residues close to the metal center on the electron donors appear to form a suitable 'pseudospecific' surface for binding to a matching negatively charged hydrophobic patch on the top of the *c* heme domain. Similar complementary surfaces have been observed for the cytochrome $c_{551}$ and nitrite reductase from *P. aeruginosa* (Nurizzo et al., 1997).

Stopped flow studies (Silvestrini et al., 1990) show that the first detectable species in the reaction under reducing conditions has the *c* heme reduced and nitric oxide bound to oxidised heme $d_1$. A mechanism for the production of this intermediate has been proposed on the basis of $H_2{}^{18}O$ exchange studies (Garber and Hollocher, 1982b; Garber and Hollocher, 1982a), which is a protonation/dehydration of nitrite bound to the ferrous heme (Figure 5.5). Two totally conserved histidine residues binding a water molecule close to the $d_1$ heme may facilitate this. The next step is proposed to be the slow loss of NO from $Fe^{3+}$, with NO known to bind less tightly to ferric than ferrous heme (Shimada and Orii, 1975). In *T. pantotropha* nitrite reductase the tyrosine ligand to the $d_1$ heme observed in the oxidized structure has been implicated in removing the NO reaction product by rebinding to the oxidized heme. However, this mechanism cannot occur for the *P. aeruginosa* enzyme where the homologous tyrosine side-chain does not bind directly to the Fe, but instead coordinates a water/hydroxide ion that is liganded to the heme (Nurizzo et al., 1997). Mutagenesis of this residue to phenylalanine results in no overall change in the

enzymatic activity or spectroscopic properties. (Cutruzzola et al., 1997). In *P. stutzeri* the N-terminal region containing this residue in the other enzymes has been deleted, inferring that a somewhat different reaction mechanism must exist for the enzyme from this species.

For the enzyme to be recycled and catalyse further reduction steps the oxidized $d_1$ heme is reduced by the *c* heme. This internal electron transfer is the rate limiting step of the reaction with a rate constant $k=1s^{-1}$(Silvestrini et al., 1990). The reduced $d_1$ heme is then ready to bind another $NO_2^-$ ion. Cytochrome $cd_1$ is inhibited by compounds that can bind at the $d_1$ heme such as CO and $CN^-$ (Silvestrini et al., 1990).

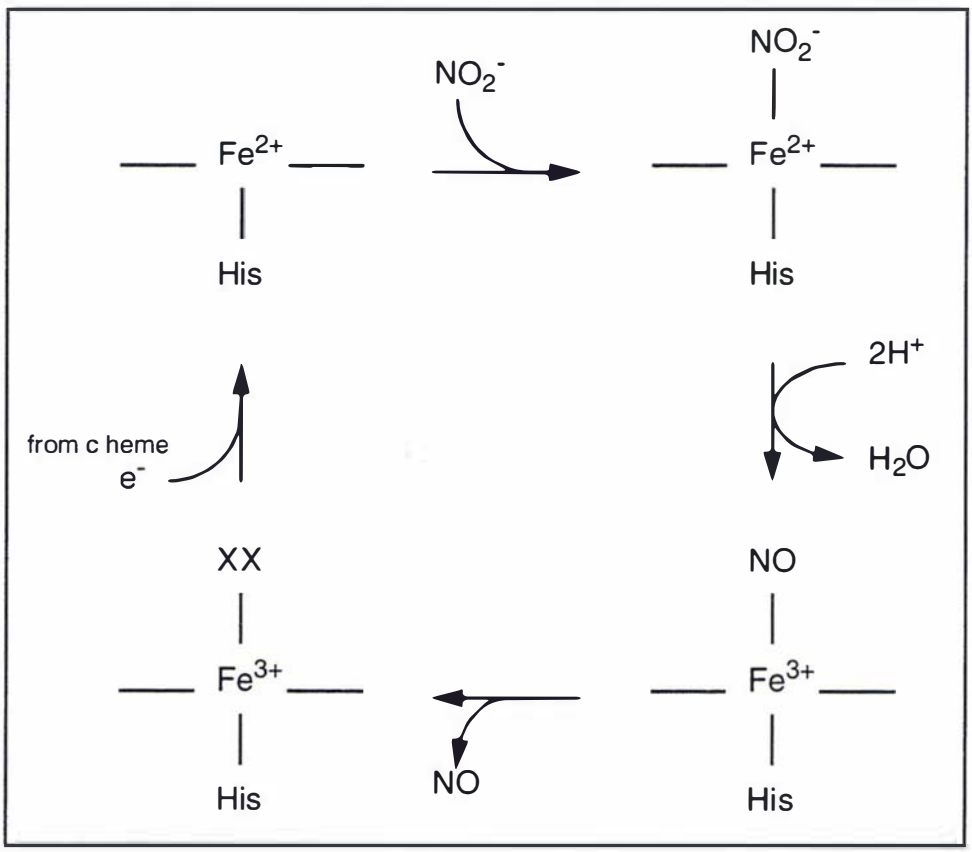



**Figure 5.5** A reaction scheme for the reduction of nitrite by cytochrome $cd_1$.

XX is the 6th heme ligand: tyrosine in *T. pantotropha* NIR or $H_2O/OH^-$ in *P. aeruginosa* NIR.

## 5.4 OTHER ENZYMES INVOLVED IN DENITRIFICATION.

### *5.4.1 Nitrate reductase*

Nitrate reductase is a heterotrimeric enzyme comprised of:

i) the 140 kDa α subunit, the site of nitrate reduction, which contains a molybdopterin guanine dinucleotide (MGD) cofactor. The molybdenum atom cycles between the Mo(IV) and Mo(VI) oxidation states via one and two electron transfer. The role of the pterin moiety is unclear. It has been suggested that this group provides greater discrimination for binding Mo over other metals than protein ligands could (Frausto da Silva and Williams, 1991), or that the delocalized electron orbitals of the pterin ring provide a lower energy electron pathway than polypeptide (Berks et al., 1995a). The method of attachment of the MGD prosthetic group to the subunit is not known, though the recent structure determination of DMSO reductase from *Rhodobacter sphaeroides* and subsequent alignment of the ligand residues indicates that nitrate reductase may have a similar coordination mode to this enzyme (Schindelin et al., 1996).

ii) the 60 kDa β subunit, which contains 4 Fe-S clusters and is presumed to be involved in electron transfer within the trimer.

iii) the γ subunit, which is integrated into the membrane and anchors the complex. The 25 kDa polypeptide accepts electrons from the membrane ubiquinol pool and transfers them to the Fe-S clusters in the β subunit. This is achieved utilizing two *b* heme groups as redox centres. Sequence analysis has lead to a proposed subunit model of 5 transmembrane helices, with four conserved histidine residues identified as potential heme ligands (Berks et al., 1995b).

### *5.4.2 Cu nitrite reductase*

There exists a second class of nitrite reductase utilized by certain bacteria for denitrification. These enzymes are copper-containing proteins and exhibit a totally different fold from the cytochrome $cd_1$ enzymes. To date no bacterium has been identified containing both types of enzyme. The distribution of the two types of nitrite reductase does not correlate with the established taxonomic classification of bacteria. The crystal structure of the enzyme has been determined from *A. cycloclastes* (Godden et al., 1991), and *A. faecalis* (Kukimoto et al., 1994). The molecule is a homotrimer, with each subunit containing a type I and type II copper centre. The type I centre is

mononuclear, coordinated to a cysteine and two histidine ligands. The copper site has an intense blue colour and is responsible for electron transfer. The type II site also contains a single copper ion bound to histidine ligands and a solvent molecule. During catalysis the bound solvent is replaced by substrate.

Each monomer consists of two β-barrel domains stacked on one another forming a layered β-sandwich. The type I centres are contained within a single monomer, whereas the type II sites are located at the subunit interfaces with each monomer contributing histidine ligands to the copper atom. The type I copper centres transfer electrons from the redox donor proteins, e.g. cupredoxins and *c*-type cytochromes, to the type II sites, where nitrite reduction occurs.

### *5.4.3 Nitric oxide reductase*

Nitric oxide reductase is comprised of two subunits. The smaller 17 kDa subunit is a cytochrome *c*-type chain, and the larger 53 kDa unit binds a *b* heme and possibly non-heme iron (Heiss et al., 1989). The genes encoding the 17 kDa (*NorC*) and 53 kDa (*NorB*) polypeptides have been cloned from *P. stutzeri* (Zumft et al., 1994). The *NorC* sequence contains the c heme binding motif Cys-X-X-Cys-His and a putative N-terminal transmembrane helix region to anchor the periplasmic domain to the cytoplasmic membrane. The *NorB* sequence shows some sequence similarity to the heme-copper oxidase family. This family incorporates a subunit that contains twelve transmembrane helices, with six conserved histidines providing ligands to two heme groups and a copper ion (Ostermeier et al., 1997). No copper has been detected in any nitric oxide reductase, but the reactive centre may contain iron instead.

### *5.4.4 Nitrous oxide reductase*

The enzyme has been purified from several species (Berks et al., 1995a) and exists as a homodimer with a monomer mass of 66-68 kDa. With the exception of the protein from *Thiobacillus denitrificans* (Hole et al., 1996), the enzyme is soluble and located in the periplasm. The enzyme binds four copper atoms per monomer, with two different types of copper sites present. A $Cu_A$ centre is thought to mediate electron transfer between an external electron donor, probably cytochrome $c_{551}$ or azurin, and the $Cu_Z$ centre, the site of nitrous oxide reduction.

## 5.5 ROLE OF STRUCTURAL STUDIES

Structural studies on cytochrome $cd_1$ nitrite reductase from *P. stutzeri* JM300 will contribute to the understanding of the structure and function of this class of heme enzyme. The structures of two nitrite reductases have been determined, from *T. pantotropha* (Fulop et al., 1995; Williams et al., 1997) and *P. aeruginosa* (Nurizzo et al., 1997). These two homologous enzymes exhibit diversity in the nature of both the *c* heme and $d_1$ heme ligands in the oxidized state, inconsistent with the hypothesis that catalytic residues are conserved within an enzyme family. The *P. stutzeri* enzyme clearly cannot share the same mode of cofactor binding as these enzymes, since the N-terminal region on which some of these coordinating residues are located is deleted. This N-terminal extension is also responsible for the majority of the interdomain contacts between the *c* heme and $d_1$ heme binding domains, so it is uncertain as to the nature of the domain interactions for *P. stutzeri* nitrite reductase. Nitrite reductase undergoes an internal electron transfer step as part of its reaction mechanism and hence examination of the *c* heme to $d_1$ heme distances and orientations would contribute to the general understanding of the nature of electron transfer between heme groups.

Chapter 6

# *P. stutzeri* Nitrite Reductase: Structure Determination

## 6.1 PURIFICATION OF NITRITE REDUCTASE

The purification of cytochrome $cd_1$ nitrite reductase from *P. stutzeri* was conducted by Dr. B. Averill and his research group at the E. C. Slater Institute, University of Amsterdam, using the method of Weeg-Aerssens (Weeg-Aerssens et al., 1991), which in turn was a modification of the protocol established by Parr and co-workers (Parr et al., 1976).

*P. stutzeri* cells were grown in soy broth containing 0.5% $KNO_3$ and 0.2% $NaHCO_3$ supplemented with small amounts of $FeSO_4$ and $CuSO_4$. A typical yield was 65g of wet cells from 15 litres of culture. The cells were then lysed by ultrasonication and the cell debris removed by centrifugation. Ammonium sulphate fractionation produced an initial crude protein preparation. Further purification was achieved by ion exchange chromatography on a DEAE-cellulose column. Detection of the nitrite reductase-containing fractions was by eye, due to the distinctive green colour of the protein. Size exclusion chromatography utilizing a Sephacryl column, followed by another ion exchange purification step resulted in nitrite reductase being eluted in two fractions. The net yield of nitrite reductase was about 40mg from the 65g of cells. The nitrite reductase is in the oxidised state in air (Weeg-Aerssens et al., 1991).

## 6.2 PROTEOLYTIC CLEAVAGE

Upon storage of nitrite reductase at 4°C over a period of two months it was observed by SDS-PAGE that degradation of the enzyme had occurred (Figure 6.1). It has been reported that the N-terminal cytochrome *c* domain and C-terminal β-propeller domain can be cleaved by subtilisin and other proteinases (Silvestrini et al., 1996), and that once separated there is no detectable reassociation suggesting that the short linker region is the only mechanism for holding the two domains together. The intensities of the two bands were very similar (as measured by eye) and the observed fragments were consistent with the larger band corresponding to the intact enzyme (~60 kDa) and the smaller species (~48 kDa) being the C-terminal β-propeller domain. There was no evidence for the small ~12 kDa fragment; a fragment this size could have been lost in the protein concentration step in which micro-centricon concentrators with membranes of molecular weight cutoffs equal to 20 kDa or 30 kDa were used.

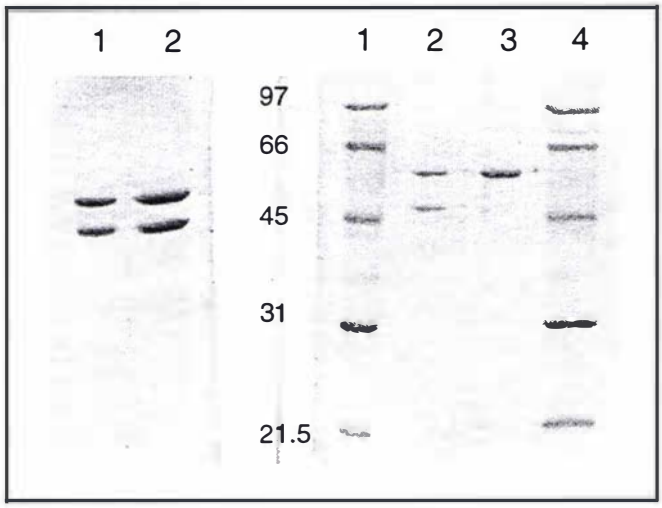



**Figure 6.1** SDS-PAGE gels of nitrite reductase samples.

left gel: two loadings of the nitrite reductase sample after prolonged storage.

right gel: lane 1. Molecular mass markers (kDa)

lane 2. Stored sample of NIR

lane 3. Fresh sample of NIR

lane 4. Molecular mass markers

In order to confirm that it was the full length molecule that had crystallized, N-terminal amino acid sequences were obtained from nitrite reductase crystals from the same crystallization experiments as those from which data was collected. The crystals were removed from their drops and washed several times with aliquots of the well reservoir solution to remove any uncrystallized protein. The crystals were then dissolved and the six N-terminal residues of the protein were sequenced by the Edman degradation reaction as run on the automated Applied Biosystems Model 476A protein sequencer. The signal for the first residue was noisy, but the remaining five residues of the sequence unambiguously corresponded to that of the intact molecule (Ala)-Ala-Pro-Asp-Met-Thr. There was no trace of any minor species present. I am most grateful to Ms J. Mudford (Massey University) for conducting the N-terminal sequencing.

## 6.3 CRYSTALLIZATION

A wide range of crystallization trials were conducted using the hanging drop method in an initial search to determine conditions suitable for crystal growth. These search experiments utilized the Kingston polymer and salt factorial screens (Kingston

et al., 1994), as well as the Jancarik and Kim screens (Jancarik and Kim, 1991). Crystals of *P. stutzeri* nitrite reductase grew at 4°C from polymer solutions, typically polyethylene glycol (PEG) or monomethyl-ether polyethylene glycol (MME-PEG) of molecular weight 4000-8000 at a concentration between 12-20%(w/v), in the pH range 6.5 to 8.5, with a protein concentration typically between 30 and 60 mg/ml (as determined by UV/visible absorbance at 280nm). The olive green crystals of oxidised nitrite reductase generally grew quickly, appearing after a period of between two to seven days. The crystals grew either as needles, or larger plates with one very thin dimension (Figure 6.2). One of the largest crystals grown had dimensions 0.1 by 0.2 by 0.6 mm. Interestingly, but unexplained at this stage, the nicest crystals generally grew around the edges of the hanging drops.

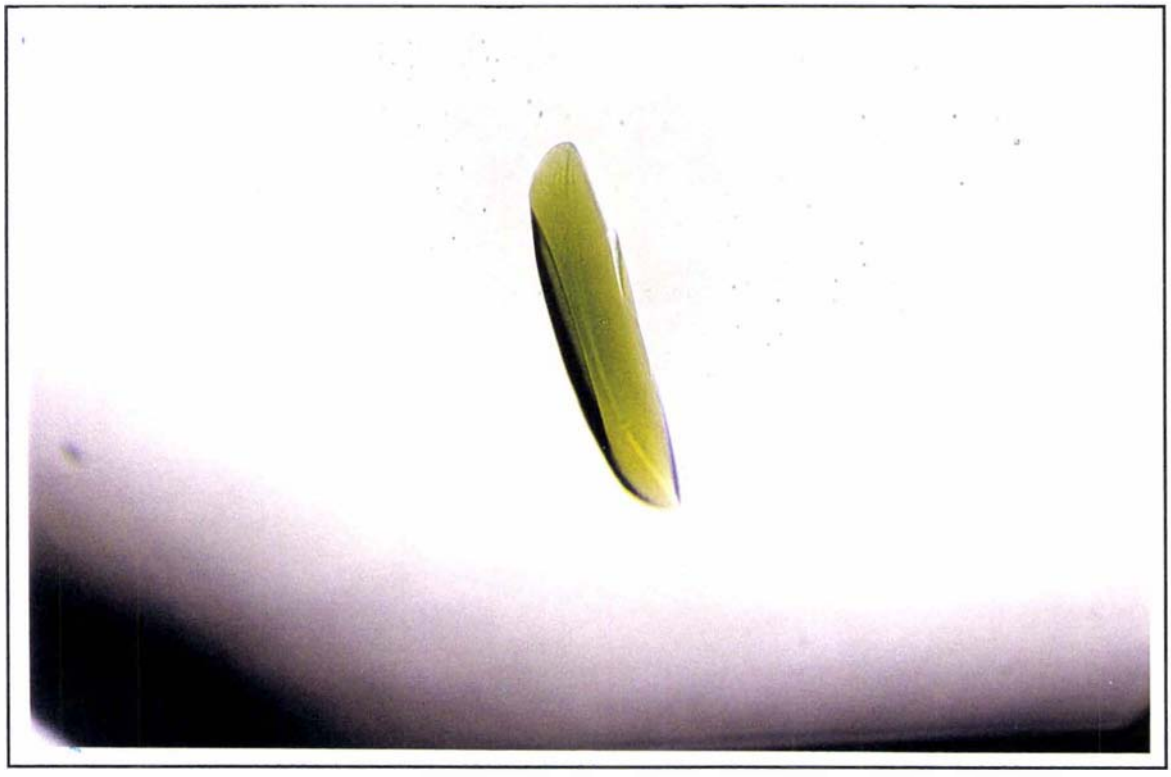


**Figure 6.2** *P. stutzeri* nitrite reductase crystal

The crystal has dimensions 0.09mm by 0.03mm by 0.5mm. The total magnification is ~100x

The crystals were generally stable in the hanging drops, but when transferred to the mother liquor well reservoir they showed signs of dissolving by cracking and frosting after several days. The crystals were also prone to cracking and fragmentation when they were manipulated for mounting.

I am very grateful to Mrs. H. M. Baker and Dr. M. Paoli (Massey University) for conducting the preliminary crystallization trials.

## 6.4 DATA COLLECTION

### *6.4.1 Room temperature data collection; Photon Factory*

#### *6.4.1.1 Data collection*

The crystal from which this data set was collected was grown in hanging drop experiments at 4°C, from 13%(w/v) MME-PEG at pH 7.2 (buffered by 0.2M MOPS), with a protein concentration of 20 mg/ml. Data were collected at the BL-6A2 synchrotron facility at the Photon Factory, Tsukuba, Japan by Prof. E. N. Baker, Dr S. A. Moore and Dr. G. B. Jameson. A Weissenberg camera is coupled to the synchrotron beamline of wavelength 1.0 Å for data collection on a Fuji imaging plate (Sakabe, 1991). Data were collected from a single crystal mounted inside a wax-sealed glass capillary with a small amount of crystallization mother liquor to prevent crystal dehydration over the course of the data collection. Small angle oscillation photographs were used to align the crystal with the needle axis parallel to the rotation axis. A total of 15 images were collected over a period of 68 minutes with an oscillation width of 12.5° and a coupling constant of 3.0°/mm. The crystal to film distance was set at 429.7 mm.

#### *6.4.1.2 Data Processing*

The auto-indexing algorithm implemented in DENZO (Otwinowski, 1993) was utilised to determine the cell dimensions, yielding a lattice consistent with a monoclinic space group, and the following unit cell dimensions:

$a$=80.6 Å, $b$=84.6 Å, $c$=105.9 Å, $\alpha$=90.0°, $\beta$=98.7°, $\gamma$=90.0°

Assuming two monomers in the asymmetric unit, the Matthews' coefficient Vm is 3.0 $Å^3Da^{-1}$, hence giving an estimated solvent content of the crystals of 59% (Matthews, 1968).

Measurement of the profile-fitted intensities was conducted using DENZO. The estimated mosaicity of the crystal was 0.4°. The measured intensities were scaled and merged employing the algorithm of Fox and Holmes (Fox and Holmes, 1966) as executed in SCALEPACK (Otwinowski, 1993). A total of 182 measurements were

rejected. No postrefinement of cell dimensions or mosaicity was conducted owing to the diffraction images being overlapped, meaning that no partial reflections could be summed between frames. The Weissenberg geometry allows very wide oscillation images to be used for data collection, resulting in the majority of reflections (90%) being fully recorded. Data processing statistics are given in (Table 6.1). Structure factor amplitudes were obtained from the measured intensities by the Bayesian statistical protocol of French and Wilson (French and Wilson, 1978), and put on to an approximate absolute scale by calculation of a Wilson plot (Wilson, 1942) using TRUNCATE from the CCP4 suite (Collaborative Computational Project, 1994). This method forces all negative intensities to be positive and inflates the weakest data which are likely to be underestimated. The Wilson estimation for the overall temperature factor was 65 $Å^2$.

**Table 6.1** Data processing statistics for NIR photon factory data set (NIR1)

| | NIR1 Data# (50-3.4 Å) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Upper resolution limit (Å) | 7.32 | 5.81 | 5.08 | 4.61 | 4.28 | 4.03 | 3.83 | 3.66 | 3.52 | 3.40 | All |
| No. of measured reflections | 3672 | 3709 | 3742 | 3685 | 3712 | 3672 | 3696 | 3627 | 3643 | 3579 | 36767 |
| No. of unique reflections | 1976 | 1946 | 1950 | 1937 | 1929 | 1899 | 1920 | 1882 | 1906 | 1865 | 19210 |
| Recorded reflections with intensity > 2σ (%) | 92.5 | 82.7 | 79.6 | 79.9 | 74.5 | 65.4 | 56.4 | 49.5 | 42.8 | 38.2 | 66.3 |
| $\langle I \rangle / \langle \sigma(I) \rangle$ | 14.8 | 8.2 | 8.1 | 7.7 | 6.9 | 5.3 | 3.9 | 3.0 | 2.5 | 2.2 | 6.4 |
| Completeness (%) | 98.7 | 99.1 | 98.6 | 98.6 | 99.2 | 99.2 | 98.2 | 97.7 | 97.3 | 96.4 | 98.3 |
| Multiplicity | 1.86 | 1.91 | 1.92 | 1.90 | 1.92 | 1.93 | 1.93 | 1.91 | 1.91 | 1.92 | 1.91 |
| R-merge* (%) | 3.7 | 9.5 | 10.4 | 9.6 | 11.5 | 14.6 | 20.5 | 26.4 | 32.0 | 35.9 | 12.3 |

#Rejected measurements are not included in the statistics

*R-merge $= \Sigma_{hkl} \Sigma_j | I_j(hkl) - \langle I(hkl) \rangle | / \Sigma_{hkl} \Sigma_j | I_j(hkl) |$

Examination of the intensities indicated that the Laue group was *2/m*. It was not possible to distinguish between the space groups $P2_1$ or P2 as no *0k0* reflections were recorded owing to the *b* axis being aligned to the rotation spindle during data collection.

Owing to the data being collected on a Weissenberg camera the anisotropy of the data was not immediately obvious from inspection of the diffraction images. The anisotropy is clearly illustrated in comparing plots of 1/resolution against the loga-

rithm of the mean native structure factor amplitudes ($<F_{obs}>$) for reflections of constant *h*, *k* and *l* respectively. All reflections of constant *h* are binned, and the mean intensity and mean 1/resolution calculated. This procedure is repeated for *k* and *l* (Figure 6.3).

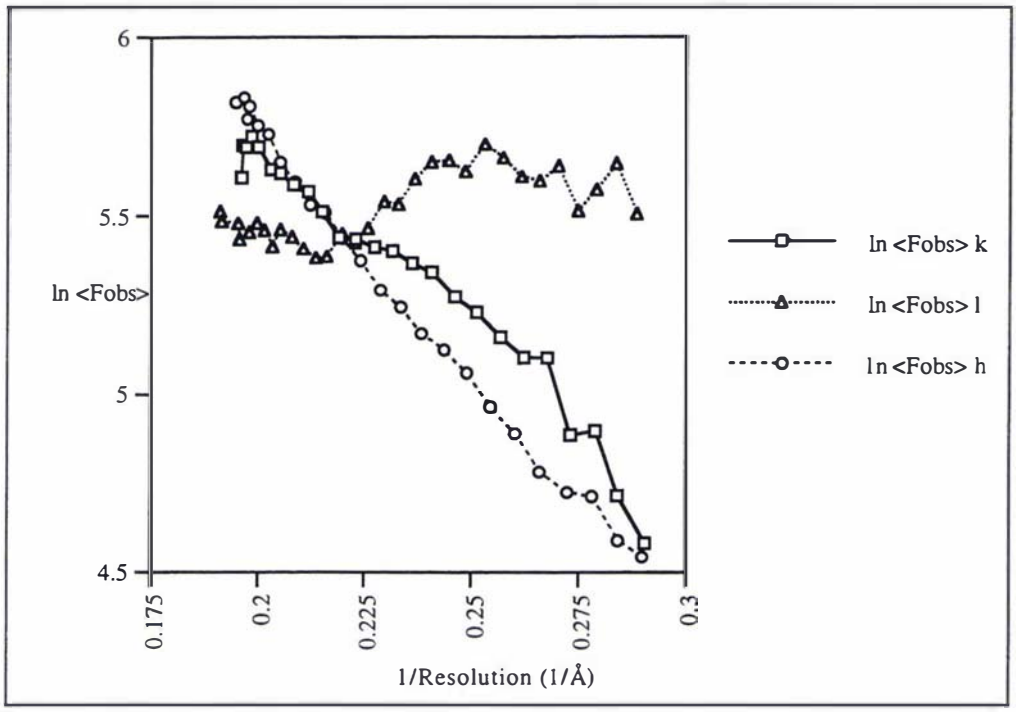



**Figure 6.3** Plot of the logarithm of the mean observed amplitude as a function of resolution for reflections of constant *h*, *k* and *l*

This figure was prepared with DATAMAN (Kleywegt and Jones, 1996).

This plot clearly shows stronger intensities at high values of *l*, indicating the increased order of the crystals in the *c** direction. The behaviour of the *h* and *k* reflections is similar, though the *h* reflections are weaker. The relative fall off of the data in the three directions, as illustrated by the three curves of this plot, contrast with those calculated for the isotropic Gower II hemoglobin data set (Figure 3.4).

A native Patterson function was calculated to investigate any non-crystallographic symmetry relationships. There were no non-origin peaks in the corresponding map, indicating the absence of any purely translational non-crystallographic symmetry.

*6.4.1.3 Self-rotation function*

The self-rotation function was calculated using the program GLRF (Tong and Rossmann, 1990). The crystal lattice is orthogonalized such that the real-space *b* axis is coincident with the cartesian y axis and $a^*$ is coincident with the x axis. The self-rotation function was calculated using spherical polar angles, defined such that $\phi$ is the angle from the x axis and $\psi$ is the angle from the z axis, with $\kappa$ being the rotation about the axis defined by $\phi$ and $\psi$. The self-rotation function was calculated in 3° steps from amplitudes within the resolution range 20.0-3.4 Å, and with magnitude greater than 200 or 8$\sigma$, as the rotation function is dominated by the large terms (Tollin and Rossmann, 1966). The top non-origin peak on the $\kappa$=180 section of the rotation function was at $\phi$=57 and $\psi$=96 (Table 6.2 and Figure 6.4). The position of this peak indicates the presence of a non-crystallographic two-fold which can be alternatively described as an axis 33° from the orientation of the crystallographic dyad along *b*, and 11° from $a^*$.

**Table 6.2** Self-rotation function calculation $\kappa$=180 section[#]

| Peak No. | Spherical Polar Angles (°) | | | Peak Height ($\sigma$) |
|---|---|---|---|---|
| | $\phi$ | $\psi$ | $\kappa$ | |
| 1 | 57.0 | 96.0 | 180 | 8.09 |
| 2 | 45.0 | 96.0 | 180 | 7.61 |
| 3 | 0.0 | 9.0 | 180 | 7.19 |
| 4 | 15.0 | 99.0 | 180 | 6.75 |

[#]Non-origin peaks shown only

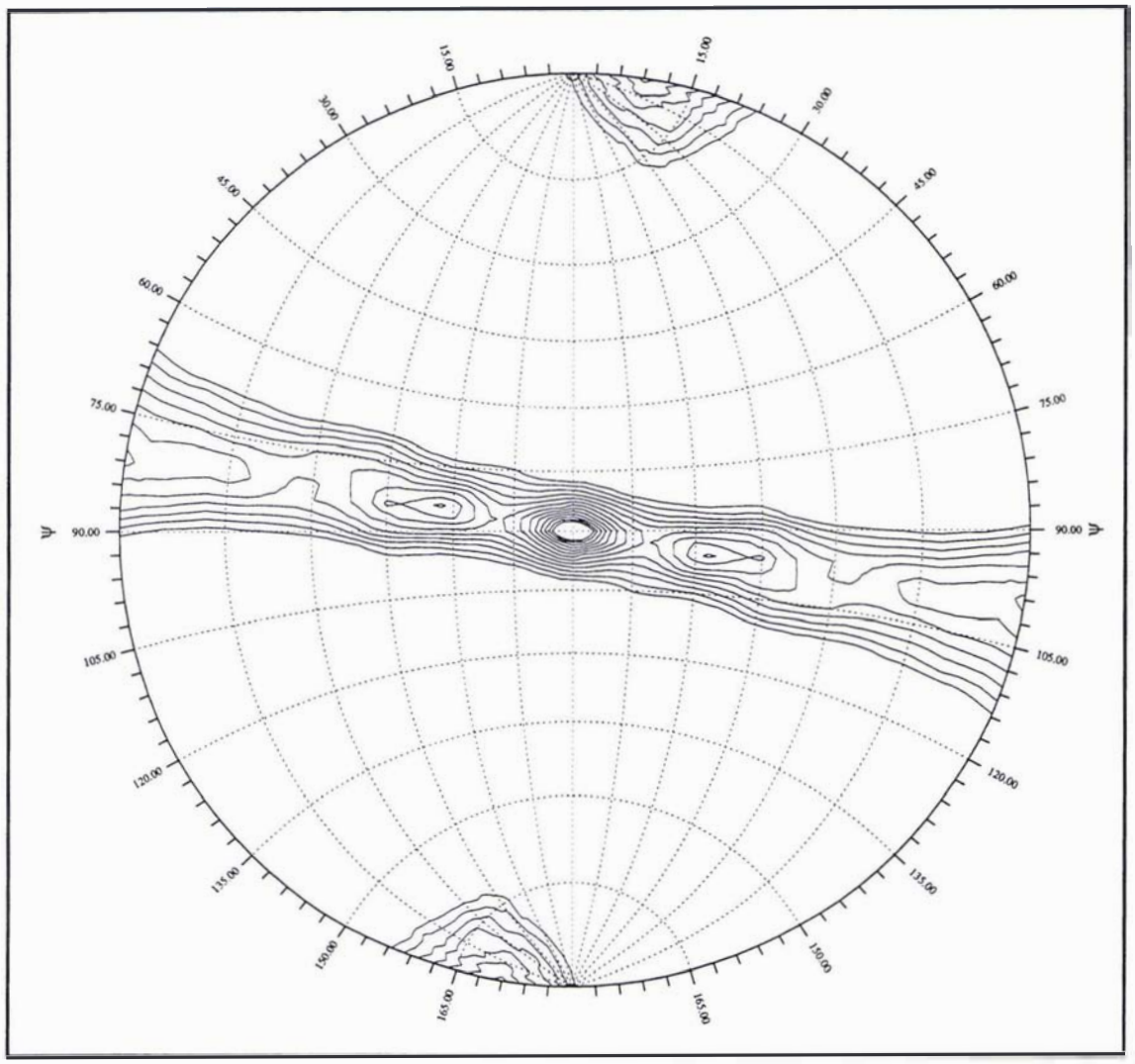



**Figure 6.4** Stereographic projection of Photon Factory data set (NIR1) self-rotation function κ=180 section

The plot of the self-rotation function has been contoured in steps of0.5σ starting at a value of 4σ above the mean value of the κ=180 section.

### *6.4.2 Room temperature data collection; Massey University.*

Further data sets were collected at Massey University using a Rigaku R-Axis IIC image plate detector with CuKα radiation of wavelength 1.54 Å from a Rigaku RU-200 rotating anode generator fitted with a 0.3mm pinhole collimator (Sato et al., 1992).

Several attempts were made to collect data from nitrite reductase crystals grown from 9-20%(w/v) PEG 4000 or PEG 6000, and 0.3-0.5M sodium acetate buffered in the pH range 6.0-6.5 (0.1M sodium cacodylate) at 4°C.

Initial X-ray experiments showed there to be a wide variability in the diffraction images obtained from different crystals, both in terms of the resolution of the data, and mosaicity of the crystals. It was not obvious from either the crystal morphology, or crystallization conditions, which crystals would be most suited for collecting X-ray data.

It was clear from these data collections that the crystals were damaged by the X-ray beam, and/or from being removed from their mother liquor, over the course of a data collection. One particular crystal initially exhibited Bragg reflections to a resolution of 2.9 Å, but after 20 hours of data collection the resolution had fallen off to 4.0 Å, with the overall data completeness at this stage only ~53%.

### *6.4.3 Cryogenic data collection; Massey University*

#### *6.4.3.1 Freezing protein crystals*

Radiation damage of protein crystals is an inherent problem in X-ray data collection experiments. The large number of oscillation images, with long exposure times, needed to collect a complete data set results in the crystal being exposed in the X-ray beam for a period of up to a few days. The X-radiation causes the formation of reactive free radicals which, over time, cause the degradation of the ordered crystal lattice. The overall effect of this is a loss of resolution and intensity in the diffraction images.

Freezing crystals at liquid nitrogen temperatures (113K) for data collection has now become an established technique enabling data to be collected from a single crystal over longer periods of time. The freezing process inhibits radiation damage, and can result in data sets with improved completeness and redundancy. The resolution limit of the data can improve, both because the higher resolution reflections are not lost over the course of the data collection, and because the ordering of the crystals can improve in some cases. On the other hand, freezing crystals often causes changes to the crystal cell axes, typically a shrinking of the cell, and can also increase the crystal mosaicity. Variability of cell dimensions of frozen protein crystals causes problems when attempting to collect isomorphous heavy atom derivatives. A com-

prehensive overview of macromolecular cryocrystallographic theory and techniques is given by Garman and Schneider (Garman and Schneider, 1997).

In order to collect a complete, and higher-resolution data set than had been achieved using crystals kept at room temperature, attempts were made to freeze the nitrite reductase crystals. In order to freeze protein crystals it is necessary to use a cryoprotectant to minimize damage to the crystal lattice by ice formation. Solutions of organic compounds used as cryoprotectants typically include various molecular weight PEGs, ethanol, glycerol, MPD, ethylene glycol and some sugars. Various concentrations of these compounds mixed with the crystal mother liquor are tested as cryoprotectants; the solution must freeze as a glass and the crystals must remain stable when transferred to the solution.

The radiation damage observed in earlier room temperature data collections suggested that freezing the nitrite reductase crystals might enable the collection of a more complete, (both overall, but particularly in the outer resolution shells) and redundant data set. Various concentrations of PEGs, glycerol and MPD were tested for suitability as cryoprotectants for the nitrite reductase crystals. Evaluation of the cryoprotectants was complicated by the varied response of the crystals when placed in the cryoprotectant solution. It was not uncommon for a crystal harvested from a particular hanging drop to partially dissolve and crack when transferred to a certain cryoprotectant, whereas another crystal from the same drop would freeze in a manner suitable for data collection. Attempts were also made to add glycerol to the crystallization conditions, initially in an attempt to improve crystal quality (Sousa, 1995) but subsequently, in an effort to grow crystals in a mother liquor that could be frozen more satisfactorily. Unfortunately adding glycerol had a deleterious effect on crystal growth.

The optimal freezing conditions comprised the crystallization mother liquor buffer and salt, 0.1M sodium cacodylate/HCl pH=6.0 and 0.4 or 0.45M sodium acetate, with the PEG 4000 concentration increased to 17-20%(w/v) and glycerol added to a concentration of 20-25%(v/v).

The procedure for freezing was to transfer the crystal from the hanging drop into a drop of the cryoprotectant which had been placed on a glass or plastic coverslip. A rayon loop of diameter typically 0.2-0.5mm, glued to a small metal base, was used to scoop up the crystal, such that the crystal sat in the loop in a small film of cryoprotectant. The loop was quickly transferred to the goniometer head situated on the

R-Axis, into the path of a liquid nitrogen-cooled air stream generated by an Oxford Cryosystem 611 (Cosier and Glazer, 1986). The cooling stream was blocked temporarily as the loop was fitted and then quickly uncovered to flash-freeze the crystal. The time the crystal spent in the cryoprotectant was kept to a minimum to reduce the chances of the crystal cracking and dissolving.

A number of crystals were tested for data collection, with two frozen crystals being ultimately utilised for data collection at Massey University using the R-Axis IIC. Initially, the measured intensities from each crystal were scaled and averaged independently (Table 6.3 and Table 6.4), but later the intensities from both crystals were scaled and merged together (Table 6.5).

*6.4.3.2 Data collection: crystal 1 (NIRf3)*

Crystal 1 was crystallized from 14%(w/v) PEG 4000 and 0.45M sodium acetate buffered at pH 6.0 using 0.1M sodium cacodylate at a temperature of 4°C. The protein concentration was 60 mg/ml, dissolved in 0.01M HEPES buffer at pH 6.8. Crystallization was achieved using the hanging drop method, with drops consisting of 2µl protein solution mixed with an equal volume of mother liquor on a siliconized glass cover slip, suspended over a mother liquor reservoir well of 800µl volume.

Data were collected from 103 oscillation images of width 1.5°. The first three frames were exposed for 50 minutes at a crystal-to-detector distance of 160 mm, with the remaining frames being exposed for the same time period but at a distance of 130 mm owing to reflections being recorded to the outer limits of the detector at the longer distance. The crystal exhibited severely anisotropic diffraction, with reflections present to a resolution of 2.4 Å in the most ordered direction, but only present to 2.9-3.0 Å over the rest of space. The diffraction images exhibited a high crystal mosaicity.

The cell dimensions obtained from DENZO were again consistent with a monoclinic space group, but had shrunk slightly in comparison to the room temperature crystals. These unit cell dimensions were:

$a$=75.6 Å, $b$=82.3 Å, $c$=102.8 Å, $\alpha$=90.0°, $\beta$=94.4°, $\gamma$=90.0°

Again assuming two monomers in the asymmetric unit, the Matthews' coefficient Vm is 2.7 $Å^3Da^{-1}$ giving an estimated solvent content of 54%.

*6.4.3.3 Data processing: crystal 1 (NIRf3)*

The intensities of the recorded reflections were measured using DENZO and subsequently scaled and averaged in SCALEPACK. A total of 25 041 measurements were rejected. This extremely high number of rejections is inflated as a consequence of the severe anisotropy in the data. In order to measure the intensity of the few high index reflections present in certain orientations it was necessary to integrate to a resolution of 2.6 Å. The integration resolution limits are applied spherically, resulting in the integration of large areas of diffraction space in which there are no observed reflection; consequently these predicted reflections are rejected as outliers. This effect is also revealed by examination of the completeness of the data as a function of resolution (see Table 6.3). Postrefinement of cell dimensions, crystal orientation parameters, and crystal mosaicity was conducted. The crystal mosaicity refined to a value of 1.1°, a higher than typical value for protein crystals (Drenth, 1994). Structure factor amplitudes were obtained from the intensities by the protocol of French and Wilson (French and Wilson, 1978), and put on to an approximately absolute scale by calculation of a Wilson plot (Wilson, 1942) using TRUNCATE from the CCP4 suite (Collaborative Computational Project, 1994). The estimated overall temperature factor was 52 $Å^2$, though this method assumes that the diffraction is isotropic, which is clearly not the case. The effective resolution of this data set was considered to be 2.8 Å on the basis of $I/\sigma(I)$ and completeness in the outermost resolution bins (Table 6.3).

**Table 6.3** Data processing statistics for frozen crystal 1 R-Axis data set (NIRf3)

| | NIRf3 Data# (40-2.6 Å) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Upper resolution limit (Å) | 5.60 | 4.44 | 3.88 | 3.53 | 3.28 | 3.08 | 2.93 | 2.80 | 2.69 | 2.60 | All |
| No. of measured reflections | 11506 | 11240 | 10565 | 10271 | 9740 | 7912 | 4490 | 3053 | 795 | 611 | 70183 |
| No. of unique reflections | 3812 | 3872 | 3841 | 3836 | 3799 | 3520 | 2580 | 1832 | 592 | 450 | 28134 |
| Recorded reflections with intensity > 2σ (%) | 96.2 | 91.2 | 81.9 | 69.1 | 56.1 | 40.3 | 32.5 | 31.7 | 26.9 | 25.3 | 64.8 |
| $\langle I \rangle / \langle \sigma(I) \rangle$ | 23.3 | 18.4 | 13.7 | 8.10 | 5.52 | 2.97 | 2.15 | 1.97 | 1.69 | 1.37 | 10.1 |
| Completeness (%) | 94.8 | 97.9 | 98.0 | 97.7 | 97.0 | 90.3 | 66.5 | 46.8 | 15.2 | 11.6 | 71.7 |
| Multiplicity | 3.0 | 2.9 | 2.8 | 2.7 | 2,6 | 2.2 | 1.7 | 1.7 | 1.3 | 1.4 | 2.5 |
| R-merge* (%) | 5.0 | 6.6 | 8.4 | 13.1 | 17.5 | 26.7 | 25.4 | 28.5 | 26.8 | 28.2 | 8.9 |

#Rejected measurements are not included in the statistics

*R-merge $= \Sigma_{hkl} \Sigma_j | I_j(hkl) - \langle I(hkl) \rangle | / \Sigma_{hkl} \Sigma_j | I_j(hkl) |$

Examination of the intensities indicated that the Laue group was *2/m* and that the *0k0* reflections obeyed the condition *0k0* absent if $k \neq 2n$, indicating that the crystals belonged to the space group $P2_1$. A subset of the measured *0k0* reflections are listed in Table 6.6

*6.4.3.4 Data collection: crystal 2 (NIRf4)*

Crystal 2 (NIRf4) was crystallized from 16%(w/v) PEG 4000 and 0.45M sodium acetate buffered at pH 6.0 by 0.1M sodium cacodylate at a temperature of 4°C. The protein concentration was 40 mg/ml dissolved in 0.01M HEPES buffer at pH 6.8.

Diffraction data were collected from eighty nine oscillation photographs of width 1.5°. The images were exposed for 35 minutes each, with the crystal to detector distance set at 120 mm. This second crystal (NIRf4) was essentially isomorphous with the first (NIRf3), with cell dimensions of:

$a$=75.8 Å, $b$=82.3 Å, $c$=102.9 Å, $\alpha$=90.0°, $\beta$=94.4°, $\gamma$=90.0°

*6.4.3.5 Data processing: crystal 2 (NIRf4)*

The measured diffraction intensities were treated in the manner described above for crystal 1 (Section 6.4.3.3), except that the data were integrated to an outer resolution of 2.8 Å. The estimate for the overall temperature factor calculated from the Wilson plot was 60 $Å^2$. Data processing statistics are given in Table 6.4.

**Table 6.4** Data processing statistics for frozen crystal 2 R-Axis data set (NIRf4)

| | NIRf4 Data# (50-2.8 Å) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Upper resolution limit (Å) | 6.03 | 4.79 | 4.18 | 3.80 | 3.53 | 3.32 | 3.15 | 3.02 | 2.90 | 2.80 | All |
| No. of measured reflections | 8466 | 8501 | 8322 | 7688 | 7462 | 7103 | 6481 | 5502 | 4478 | 3396 | 67399 |
| No. of unique reflections | 2983 | 3001 | 3024 | 2999 | 3044 | 2994 | 2928 | 2663 | 2296 | 1800 | 27732 |
| Recorded reflections with intensity > 2σ (%) | 96.1 | 92.4 | 87.8 | 78.4 | 71.5 | 62.4 | 46.9 | 38.9 | 32.2 | 33.6 | 66.9 |
| <I>/<σ(I)> | 23.2 | 19.0 | 16.8 | 11.5 | 7.83 | 6.02 | 3.65 | 2.67 | 2.42 | 2.08 | 10.1 |
| Completeness (%) | 91.9 | 94.6 | 95.5 | 95.8 | 96.4 | 96.1 | 93.2 | 84.8 | 73.1 | 58.3 | 88.0 |
| Multiplicity | 2.8 | 2.8 | 2.8 | 2.6 | 2.5 | 2.4 | 2.2 | 2.1 | 2.0 | 1.9 | 2.4 |
| R-merge* (%) | 3.8 | 9.2 | 10.8 | 10.8 | 11.7 | 14.2 | 18.8 | 22.9 | 28.7 | 33.2 | 9.4 |

#Rejected measurements are not included in the statistics

*R-merge $= \Sigma_{hkl} \Sigma_j | I_j(hkl) - \langle I(hkl) \rangle | / \Sigma_{hkl} \Sigma_j | I_j(hkl) |$

To obtain a more complete data set, with a higher redundancy, the two sets of intensities were combined for scaling and averaging in the manner described previously (Section 6.4.3.3). A total of 137 582 reflections were combined, yielding a data set (NIRc) with an overall multiplicity of 4.2 (for data in the range 50-2.8 Å). A summary of the processing statistics is given in Table 6.5. The overall temperature factor as estimated from a Wilson plot was 70 $Å^2$, a value higher than that calculated for the individual data sets. This may be due to slight non-isomorphism between the two crystals, with the post-refined cell dimensions differing by 0.2 Å along *a*, and 0.1 Å along *c*, though a recent treatment of this problem suggests that systematic differences between crystals are generally small in comparison to random errors and that multiple crystal data sets with increased multiplicity will result in more accurate reduced structure factors (Diederichs and Karplus, 1997).

**Table 6.5** Data processing statistics for the combined frozen R-Axis data sets (NIRc)

| | NIRc Data# (50-2.6 Å) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Upper resolution limit (Å) | 5.60 | 4.44 | 3.88 | 3.53 | 3.28 | 3.08 | 2.93 | 2.80 | 2.69 | 2.60 | All |
| No. of unique reflections | 4017 | 3954 | 3916 | 3923 | 3905 | 3753 | 3274 | 2584 | 623 | 453 | 30402 |
| Recorded reflections with intensity > 2σ (%) | 97.0 | 93.3 | 86.3 | 77.6 | 67.3 | 49.5 | 39.2 | 23.0 | 26.3 | 24.9 | 91.2 |
| $\langle I \rangle / \langle \sigma(I) \rangle$ | 31.2 | 26.7 | 20.3 | 12.2 | 8.44 | 4.44 | 2.94 | 2.48 | 1.67 | 1.36 | 14.4 |
| Completeness (%) | 99.9 | 100.0 | 99.9 | 99.9 | 99.7 | 96.2 | 84.4 | 66.0 | 15.9 | 11.7 | 77.5 |
| R-merge* (%) | 5.9 | 7.5 | 9.7 | 14.6 | 19.5 | 29.1 | 33.0 | 34.0 | 29.0 | 28.4 | 10.2 |

#Rejected measurements are not included in the statistics

*R-merge $= \Sigma_{hkl} \Sigma_j | I_j(hkl) - \langle I(hkl) \rangle | / \Sigma_{hkl} \Sigma_j | I_j(hkl) |$

A subset of the measured *0k0* reflections is listed in Table 6.6.

**Table 6.6** Subset of *0k0* intensities for data set NIRc

| *h* | *k* | *l* | Intensity | Sigma | I/Sigma |
|---|---|---|---|---|---|
| 0 | 3 | 0 | 1.7 | 4.6 | 0.4 |
| 0 | 4 | 0 | 28891 | 1155 | 25 |
| 0 | 5 | 0 | 4.7 | 5.7 | 0.8 |
| 0 | 6 | 0 | 8092.7 | 376.0 | 21.5 |
| 0 | 7 | 0 | -2.6 | 8.0 | -0.3 |
| 0 | 8 | 0 | 1942 | 77.0 | 25.2 |
| 0 | 20 | 0 | 2206.0 | 82.6 | 26.7 |
| 0 | 21 | 0 | -28.4 | 32.2 | -0.9 |
| 0 | 22 | 0 | 897.2 | 59.5 | 15.1 |
| 0 | 23 | 0 | 23.3 | 52.8 | 0.4 |
| 0 | 24 | 0 | 166.3 | 33.2 | 5.0 |
| 0 | 25 | 0 | 48.4 | 29.8 | 1.6 |
| 0 | 26 | 0 | 86.1 | 39.1 | 2.2 |

*6.4.3.6 Self-rotation function: combined frozen data sets.*

A self-rotation function was calculated with GLRF, using the conventions described previously (Section 6.4.1.3). Data in the resolution range 10 to 3.5 Å, with intensities of greater than 4σ, or with magnitude greater than 100.0 were used. The results of the calculation are listed in Table 6.7 and represented as Figure 6.5. The top non-origin peak on the κ=180 section was at φ=60 and ψ=90, indicating the presence of

an NCS dyad axis perpendicular to z and 30° from y (which is aligned with the *b* axis). This represents a small shift from that found for the room temperature data set.

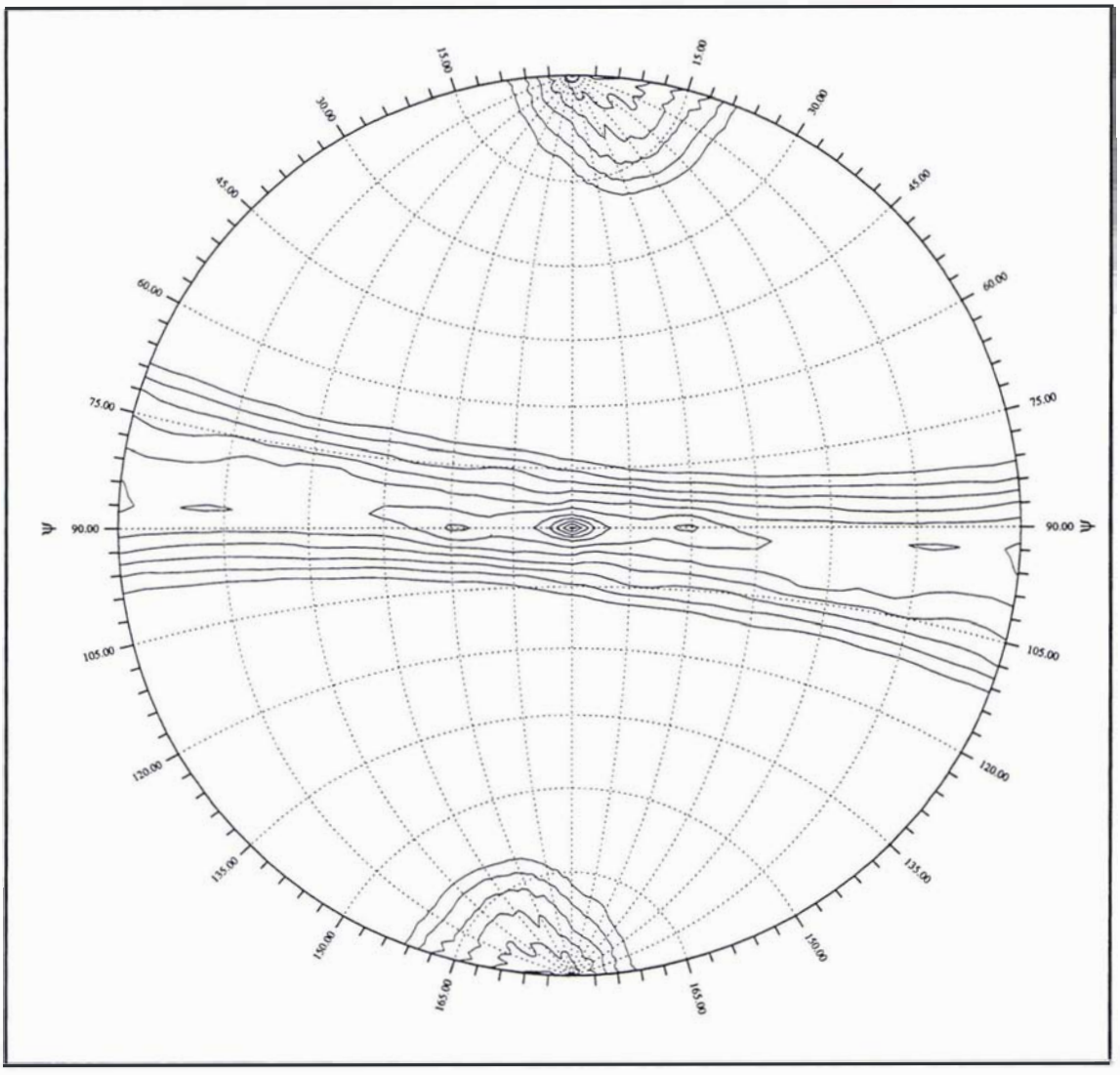



**Figure 6.5** Stereographic projection of combined frozen data sets (NIRc) self-rotation function κ=180 section.

The plot of the self-rotation function has been contoured in steps of 0.5σ, starting at a value of 1.5σ above the mean value of the κ=180 section.

**Table 6.7** Self-rotation function calculation κ=180 section[#]

| Peak No. | Spherical Polar Angles (°) | | | Peak Height (σ) |
|---|---|---|---|---|
| | ϕ | ψ | κ | |
| 1 | 60.0 | 90.0 | 180 | 4.62 |
| 2 | 48.0 | 93.0 | 180 | 4.14 |
| 3 | 0.0 | 93.0 | 180 | 4.10 |
| 4 | 12.0 | 93.0 | 180 | 4.04 |

[#]Non origin peaks shown only

Calculation of the self-rotation function (as above) for values of κ from 0 to 180 reveals the presence of a large peak at ϕ=0, ψ=0 and κ=60, indicating the presence of a 60° rotation about the z axis. This NCS operation is consistent with an NCS two-fold oriented 30° from the *b* axis lying in the *ab* plane. The presence of a non-crystallographic symmetry axis in a non-general orientation is not uncommon for monoclinic space groups (Wang and Janin, 1993). This peak was the only feature on the κ=60 section.

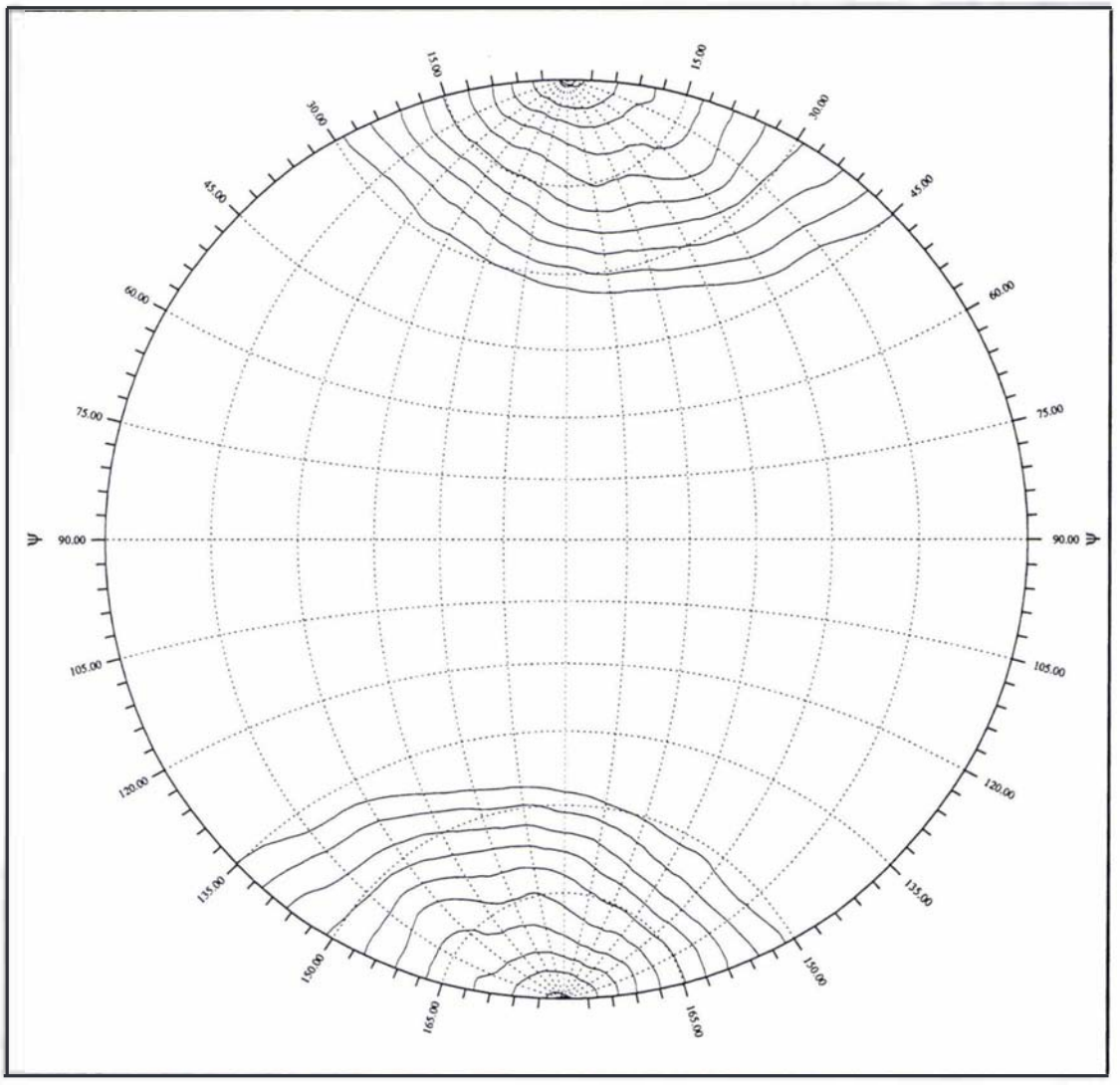



**Figure 6.6** Stereographic projection of combined frozen data sets (NIRc) self-rotation function κ=60 section.

The plot of the self-rotation function has been contoured in steps of 0.5σ, starting at a value of 1.5σ above the mean value of the κ=60 section.

### *6.4.3.7 Anisotropy in the frozen crystal data sets*

From examination of the oscillation images it was clear that the diffraction from the frozen nitrite reductase crystals was appreciably anisotropic (Figure 6.7). Anisotropic diffraction is a function of the packing inside the crystal lattice. If crystal contacts are sparse in one direction this can result in disorder of the lattice in that orientation, hence causing a decrease in the resolution of the diffracted data that is uneven over reciprocal space. This effect is not uncommon for protein crystals. Some examples

are described in the introduction to the Sheriff and Hendrickson treatment of anisotropic diffraction from protein crystals (Sheriff and Hendrickson, 1987), as well as in more recent publications (Garcia et al., 1998, Tesmer et al., 1997, Garboczi et al., 1996, Lu et al., 1995). Examination of the reduced structure factor amplitudes from the frozen nitrite reductase crystals revealed that *a** was the most disordered direction with the absence of any reflections with high *h* indices, whereas *c** was the most well ordered direction, with *b** intermediate.

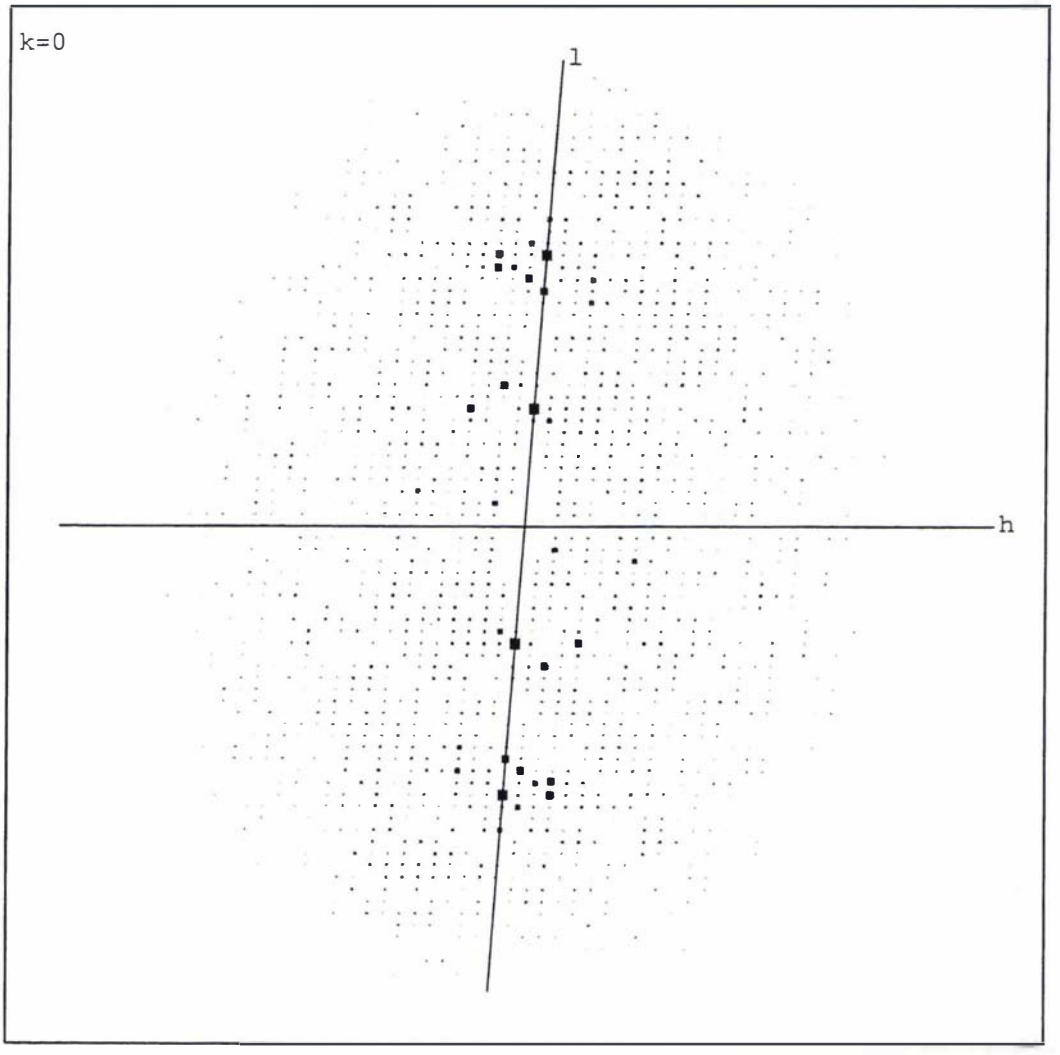



**Figure 6.7** Representation of the *h0l* section of the frozen nitrite reductase reciprocal lattice

This figure was prepared with HKLVIEW (Collaborative Computational Project, 1994)

The anisotropy of the data is clearly illustrated by plots of 1/resolution against the logarithm of the mean native structure factor amplitudes for reflections of constant *h*, *k* and *l* respectively (Figure 6.8).

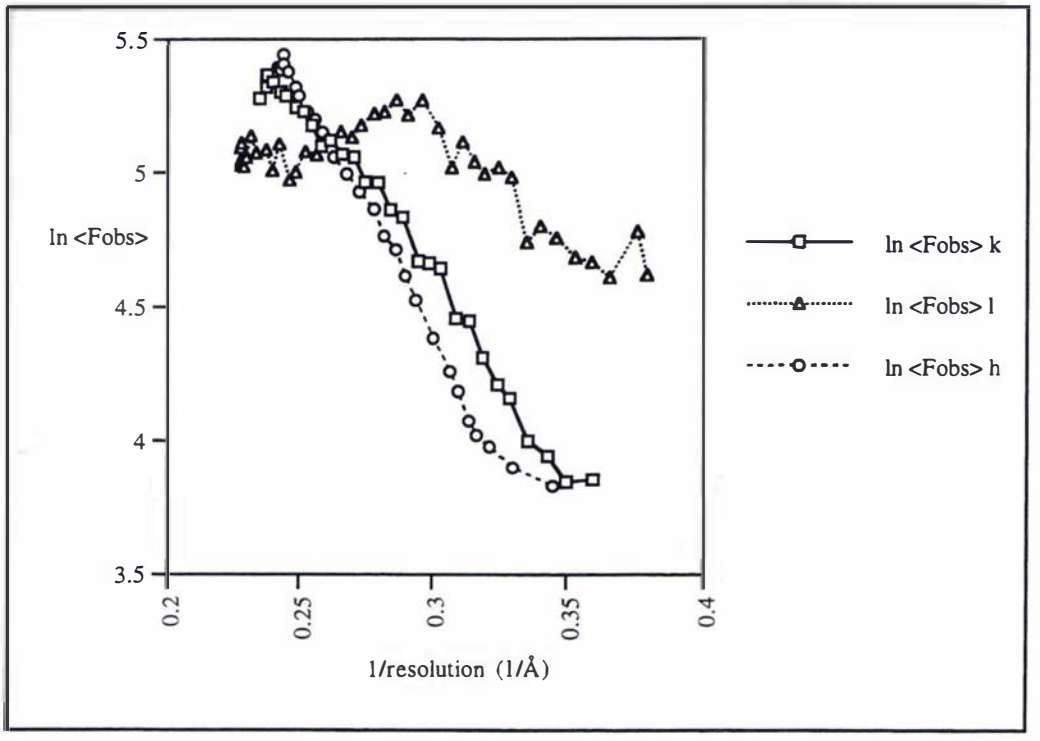



**Figure 6.8** Plot of the logarithm of the mean observed amplitude as a function of resolution for reflections of constant *h*, *k* and *l*

This figure was prepared with DATAMAN (Kleywegt and Jones, 1996).

This plot clearly shows the stronger intensities with index *l*, indicating the increased order of the crystals in the $c^*$ direction. The behaviour of the *h* and *k* reflections is similar, though the *h*reflections are weaker. Similar results are apparent for the room temperature data (Section 6.4.1.2), indicating that freezing the crystals did not alter the anisotropy of the data, though the mosaicity did increase.

## 6.5 MOLECULAR REPLACEMENT

Owing to the availability of an appropriate search model, the molecular replacement method was used to solve the structure of *P. stutzeri* nitrite reductase.

### 6.5.1 The search model

An initial search model (model1) was prepared using the coordinates of cytochrome $cd_1$ nitrite reductase from *Thiosphaera pantotropha*, refined at high resolution (Fulop et al., 1995). The model used was a monomer comprising both the *c* heme (N-terminal) and $d_1$ heme-binding (C-terminal) domains as well as the heme groups themselves. All water molecules were removed and the temperature factors were set to 50.0 $Å^2$. A more sophisticated treatment of the model B-factors, that has not been explored further, might have been to scale the temperature factors so that the average B-factor was 50.0 $Å^2$ rather than setting them all to the one value. From examination of the refined model and alignments of the *T. pantotropha* and *P. stutzeri* sequences, it was possible to estimate where any insertions and deletions between the two polypeptides occurred. Any insertions in the known structure were removed from the model, along with the surrounding residues, with the number omitted dependant on the level of sequence identity in that region. Side-chain substitutions between the two sequences were treated in the following manner: aspartate/asparagine, glutamate/glutamine and valine/threonine substitutions were considered to be identical and left unchanged. Some side-chains in the *T. pantotropha* nitrite reductase structure were truncated to match the *P. stutzeri* sequence, where they retained similar chemical properties, i.e. tyrosine → phenylalanine, isoleucine → valine, threonine → serine, arginine → lysine and alanine → glycine. All other amino acid positions with side-chain substitutions were truncated to alanine, with the exception of glycine which were left unaltered.

A second model (model 2) was built in the same manner as described above from the coordinates of *P. aeruginosa* nitrite reductase, refined at 2.8 Å (now refined at 2.2 Å, Nurizzo et al., 1997).

A third model (model 3) was prepared from the *T. pantotropha* nitrite reductase coordinates in an identical manner to the first, except that the N-terminal *c* heme-binding domain and *c* heme were omitted.

### 6.5.2 Rotation and translation functions

Reciprocal space rotation and translation functions were calculated using AMORE (Navaza, 1994) as implemented in the CCP4 suite. These functions were calculated using reflections between 8 and 4 Å resolution, from the frozen data set (NIRf3). For

the rotation function calculation the outer Patterson integration radius was 30.0 Å, with no inner cut-off. The results of the rotation function are given in Table 6.8.

**Table 6.8** Model 1 rotation function solutions

| Eulerian angles (°) | | | Peak Height |
|---|---|---|---|
| α | β | γ | (σ) |
| 76.5 | 46.09 | 221.12 | 17.8 |
| 15.26 | 42.77 | 225.26 | 14.4 |
| 307.19 | 77.39 | 239.03 | 8.3 |
| 177.80 | 129.33 | 237.22 | 7.7 |
| 335.0 | 80.89 | 241.37 | 7.7 |

Eulerian angles (α, β, γ) are defined such that α is a rotation of the coordinate system about z, β is a rotation about the new y axis, and γ is a rotation about the new z direction. The output angles are those required to orient the model after it has had its centre of mass placed at the origin and been aligned by its principal axes of inertia to the cell axes.

Two clear peaks above the noise level could be resolved in the rotation function. These two solutions were separated by an approximately 60° rotation, which is consistent with the orientation of the NCS dyad as observed in the self-rotation function (Section 6.4.3.6). The top rotation function solutions were then used to orient the model and calculate the translation function to position the first monomer. The top two translation function solutions are listed in Table 6.9 for each of the three highest rotation solutions.

**Table 6.9** $P2_1$ model 1 translation function solutions: first molecule

| Rotation Function Solutions (Eulerian angles) (°) | | | Translation Function Solutions (Fractional Coordinates) | | | Correlation Coefficient | R-factor |
|---|---|---|---|---|---|---|---|
| α | β | γ | x | y | z | | |
| 76.5 | 46.09 | 221.12 | 0.3125 | 0 | 0.0648 | 0.258 | 0.552 |
| 76.5 | 46.09 | 221.12 | 0.2750 | 0 | 0.1759 | 0.218 | 0.563 |
| 15.26 | 42.77 | 225.26 | 0.1375 | 0 | 0.3704 | 0.200 | 0.568 |
| 15.26 | 42.77 | 225.26 | 0.3875 | 0 | 0.1667 | 0.188 | 0.578 |
| 307.19 | 77.39 | 239.03 | 0.2125 | 0 | 0.2315 | 0.140 | 0.586 |
| 307.19 | 77.39 | 239.03 | 0.0500 | 0 | 0.1944 | 0.137 | 0.589 |

The top solution from the translation function was fixed in the asymmetric unit, and a search conducted for the position of the second molecule. The results of this calculation are listed in Table 6.10.

**Table 6.10** $P2_1$ model 1 translation function solutions: second molecule

| Rotation Function Solutions (Eulerian angles) (°) | | | Translation Function Solutions (Fractional Coordinates) | | | Correlation Coefficient | R-factor |
|---|---|---|---|---|---|---|---|
| α | β | γ | x | y | z | | |
| 15.26 | 42.77 | 225.26 | 0.6340 | 0.5321 | 0.3717 | 0.346 | 0.524 |
| 15.26 | 42.77 | 225.26 | 0.5602 | 0.5595 | 0.8890 | 0.240 | 0.563 |
| 307.19 | 77.39 | 239.03 | 0.5540 | 0.2351 | 0.7071 | 0.227 | 0.566 |
| 307.19 | 77.39 | 239.03 | 0.4441 | 0.1133 | 0.0732 | 0.221 | 0.566 |

The search model was oriented and positioned by the two molecular replacement solutions and subjected to fifteen cycles of rigid-body fitting in AMORE, resulting in a model for the nitrite reductase dimer with a correlation coefficient of 0.468 and an R-factor of 0.499 (for data in the range 8.0 -4.0 Å).

Using the *P. aeruginosa* nitrite reductase monomer (model 2) coordinates as a search model yielded the same molecular replacement solution (results not shown), but with a lower correlation coefficient (0.429) and higher R-factor (0.509).

The translation function calculations were also carried out for the space group P2, with the results for placement of the second monomer given in Table 6.11.

**Table 6.11** Model 1 P2 translation function solutions second molecule

| Rotation Function Solutions (Eulerian angles) (°) | | | Translation Function Solutions (Fractional Coordinates) | | | Correlation Coefficient | R-factor |
|---|---|---|---|---|---|---|---|
| α | β | γ | x | y | z | | |
| 15.26 | 42.77 | 225.26 | 0.5051 | 0.0265 | 0.4243 | 0.269 | 0.549 |
| 15.26 | 42.77 | 225.26 | 0.3891 | 0.1809 | 0.9207 | 0.233 | 0.565 |
| 307.19 | 77.39 | 239.03 | 0.4522 | 0.0306 | 0.6840 | 0.215 | 0.567 |
| 307.19 | 77.39 | 239.03 | 0.4819 | 0.7728 | 0.0351 | 0.211 | 0.573 |

Fifteen cycles of rigid-body fitting in AMORE resulted in the best P2 solution having a correlation coefficient of 0.379 and an R-factor of 0.534 (for data 8.0-4.0 Å). These results were consistent with the systematic absences seen for the *0k0* reflections, and confirmed that the space group was $P2_1$.

### *6.5.3 Validation of the molecular replacement solution*

This solution was considered correct for several reasons. The correlation coefficient and R-factor values are in the region normally found for reasonable molecular replacement solutions, and more importantly there is good discrimination of the solution above the noise peaks. The orientation of the two molecules placed in the asymmetric unit is consistent with the information contained in the self-rotation function. Using a monomer as the search model meant that no prior assumptions of how the nitrite reductase dimer was formed were imposed on the molecular replacement calculations. The two independently placed monomers from the molecular replacement solution resulted in the formation of a dimer that was essentially identical to that found for *P. aeruginosa* and *T. pantotropha* nitrite reductases. When the crystallographic symmetry is applied to the asymmetric unit the molecules pack to form a sensible lattice, with no steric clashes (Figure 7.15).

Rigid-body refinement of the solution was conducted in TNT (Tronrud et al., 1987) with data of resolution 20-2.8 Å, initially refining with the four domains as rigid units and finally with the molecule divided into fragments where the chain breaks fell naturally in the model. The free R-factor dropped from 0.542 to 0.524 with a corresponding decrease in the conventional R-factor (0.538 to 0.525). From inspection of SIGMAA-weighted (Read, 1986) electron density maps calculated from the model it was immediately apparent that while the C-terminal β-propeller domains were correctly placed, both the N-terminal domains of the dimer appeared to be in the wrong orientation and/or position in the cell.

The molecular replacement calculations were repeated using the same parameters as described previously, but with search model 3, in order to more accurately determine the position of the C-terminal domains. The same molecular replacement solutions were obtained, but with an improved signal-to-noise ratio. The correlation coefficient (cc) and R-factor (R) after placing the second molecule were 0.417 and 0.499. Rigid-body fitting improved the model further (cc=0.525 and R=0.478).

Rigid-body refinement in TNT following the same protocol as for the whole monomer resulted in the free R-factor decreasing to 0.455 with a conventional R-factor of 0.431. The omission of the two N-terminal domains from the search model thus resulted in a drop of 7% in the free R-value after rigid-body refinement.

This model was then used to calculate an initial phase set and provide a starting point for the location of the N-terminal domains and refinement of the structure.

## 6.6 LOCATION OF THE N-TERMINAL DOMAINS

### *6.6.1 Modelling by homology*

From the molecular replacement calculations it was clear that the N-terminal domains could not be located in a similar orientation and/or position to that found for the *P. aeruginosa* and *T. pantotropha* nitrite reductases. Attempts were made to place just the *c* heme group and the strictly conserved heme-binding motif (residues 21-35 for the *P. stutzeri* nitrite reductase sequence) in a position corresponding to that in the other two enzymes. Inspection of electron density maps calculated from the coordinates of the β-propeller domain and these fragments indicated that the position of this core part of the N-terminal domain structure was not conserved. From the initial maps phased with the β-propeller domain coordinates, little density could be seen that could indicate the position of the N-terminal domain (Figure 6.9).

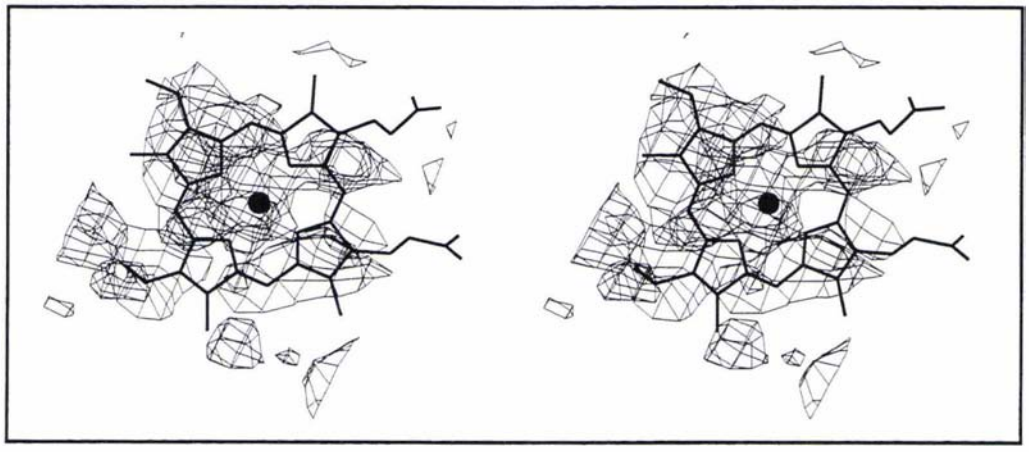


**Figure 6.9** Stereo view of the electron density in the region of the N-terminal domain phased from the molecular replacement solution.

The coordinates of the current *c* heme group are overlaid on the map. The $2m|F_o| - D|F_c|$ map is contoured at 1 sigma. This figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

### 6.6.2 Anomalous signal

Attempts were made to utilize the anomalous scattering of the iron atoms in order to position the *c* hemes. The frozen and room temperature data sets were reprocessed without merging of Friedel pairs. Examination of the SCALEPACK $\chi^2$ value, which in this case represents the weighted ratio of the difference of the anomalous pairs, indicated that there was no detectable anomalous signal. Calculation of $|F^+|-|F^-|$ maps with phases computed from the β-propeller domains with the $d_1$ heme iron omitted yielded no information about the position of the *c* heme iron. In addition there was no peak present for the $d_1$ heme iron which was known to be correctly positioned, confirming that there was no useful anomalous signal in either data set.

### 6.6.3 Molecular replacement methods

Owing to the unusual ligands for the *c* heme in the *T. pantotropha* nitrite reductase structure and the greater sequence identity between the N-terminal domains of the *P. stutzeri* and *P. aeruginosa* nitrite reductases (section 5.3.4 on page 99), a single N-terminal domain from the latter enzyme was used as a search model. A monomer was used, as it could not be safely assumed that the NCS relationship of the two N-terminal domains would necessarily correspond to that of the β-propeller domains. Two search models were utilised; (i) the whole N-terminal domain and *c* heme with the large N-terminal insertion removed and side-chains modified in the manner described for the β-propeller domain models, and (ii) just the conserved *c* heme-binding motif (15 residues) and *c* heme. Rotation functions were calculated using the programs AMORE (Navaza, 1994) and X-PLOR (Brunger, 1990), however no solution was apparent. Translation functions were calculated with the top rotation solution peaks and yielded no discernible position for the search model. The failure of these molecular replacement calculations was not surprising considering that one N-terminal domain (99 residues) comprises less than 10% of the total scattering of the asymmetric unit.

A more promising approach appeared to be the method of Zhang and Matthews (Zhang and Matthews, 1994), in which information from an already determined part of the structure is utilised to help orient and position the unknown part. In the rotation function calculation the Patterson function of the known fragment is 'subtracted' from the native Patterson and thus the remaining observations now more closely resemble the remaining part of the structure. This procedure eliminates peaks that are not relevant to the required solution. This method can only be carried out in Patterson

space and requires an accurate scale factor between the observed intensities ($I_o$) and those calculated from the partial known model ($I_a$). The scale factor ($k$) is estimated from Wilson statistics and knowledge of the fraction ($f$) of the asymmetric unit that the partial structure comprises that is to be subtracted.

$$k = f\sum_h I_o(h) / \sum_h I_a(h) \qquad \text{Eq. 6.1}$$

The modified correlation form of the rotation function, including the subtracted component, becomes:

$$C(R) = \langle (P_o - kP_a) | RP_b \rangle \times [\langle (P_o - kP_a) | (P_o - kP_a) \rangle \langle P_b | P_b \rangle]^{-1/2} \qquad \text{Eq. 6.2}$$

where $P_a$ is the Patterson function calculated from the known model fragment and $P_b$ is calculated from the search model. Rotation functions of this form were calculated using the N-terminal domain search models described above, and the correctly positioned C-terminal β-propeller domains, after limited rebuilding and refinement, as the known structural fragment. However no discernible solution for the orientation of the N-terminal domain was found. Translation functions that incorporate known structural information in an additive manner were calculated using the top rotation solution peaks, but gave no indication of the position of the N-terminal domain in the cell. The failure of this method is probably due to the difficulty in determining an appropriate scale for the native data and those calculated from the known fragment. The moderate resolution of the data results in Wilson statistics being barely justified. This factor, coupled with the anisotropy present in the data, results in an inadequate scaling function and hence a poor estimate of the Patterson function calculated from the known partial structure that is to be subtracted. The problems in determining the orientation of the N-terminal domain were compounded by the apparent disorder that was later found to be displayed by this part of the structure.

An alternative approach that allows the incorporation of known structural information into molecular replacement calculations is that of Fujinaga and Read, as implemented in the program BRUTE (Fujinaga and Read, 1987). A linear correlation coefficient is used to determine the position of a correctly oriented molecule or fragment within the cell. The orientation of the model can be adjusted during the translation search, though only to a small degree, otherwise the resulting six dimensional search becomes computationally unfeasible. This method does allow for the inclusion of a known partial model which is fixed in the cell, and its contribution to the cal-

culated structure factors is added to those of the moving search model. The initial orientation and position of the N-terminal search models (as described above) was assumed to be the same as found for the *T. pantotropha* and *P. aeruginosa* nitrite reductases. This procedure tested whether a small rigid-body shift of the N-terminal domain had occurred within the *P. stutzeri* monomer. The search model was rotated about its initial position by ±11.5° in 3° steps around each of the three axis, while being translated by ±4 Å, in 1 Å steps, in the x, y and z cartesian directions. Even this limited search took a prohibitive amount of computer time. No orientation and position for the shifted search model gave a discernible solution. These calculations appeared to confirm what was evident from the electron density maps phased just from the β-propeller domains, that the N-terminal domain of the *P. stutzeri* nitrite reductase had undergone a rigid-body shift relative to the β-propeller domain, compared to the homologous enzymes from *T. pantotropha* and *P. aeruginosa*, in addition to any tertiary conformational changes.

### *6.6.4 Bootstrapping and dummy atom procedures*

Having been unable to place a fragment of the N-terminal domain by homology with the other nitrite reductases, the remaining option was to build into what features could be perceived in the electron density maps. These efforts were conducted on maps that were slowly improving due to continuing refinement and rebuilding of the β-propeller domains. From the sparse features in the maps in the region of the N-terminus of one of the monomers, it was clear that a significant rigid-body movement of the N-terminal domains had occurred. In order to facilitate the rapid building into these poor maps a dummy atom procedure, utilised with some success previously by Dr. R. Kingston (Kingston, 1996), was employed. Dummy atoms were built into the maps at the estimated positions of the Cα atoms for one of the N-terminal domains. Because no stereochemical or directional information about the protein chain were required for this building the procedure is a very fast process. Maps were calculated employing the globic scattering factors proposed by Guo et al (Guo et al., 1995) for these dummy atoms. These scattering factors were incorporated into TNT for computation of the Fourier maps and limited refinement. During refinement, non-contact restraints were the only stereochemical information enforced on the dummy atoms. This protocol was extremely useful for quickly establishing the connectivity of isolated density features in the maps. However the phase error from these scattering factors is 80° at 3.0 Å, resulting in the maps taking on a very globular nature in these regions, obscuring any information in the maps about the position of the peptide carbonyls or amino acid side-chains. The dummy atoms that

remained in connected density were then replaced with a polyalanine model. However, after several rounds of iterative rebuilding and refinement, little new information was apparent in the new maps that could indicate where the remainder of the structure was located, or what adjustments were required to improve the partial model present. For example, it was not possible to determine the position of the *c* heme group from the maps at this stage. This led to this particular approach being abandoned at this stage, with the model having an R-factor of 0.330 and a free R-factor of 0.387 (with an anisotropic correction applied, see Section 6.7.5). In hindsight (see Figure 6.10) it is clear that substantial errors had been incorporated into the model, and with no independent phase information and a native data set of questionable quality it would have been difficult to proceed from this point using the methods discussed above.

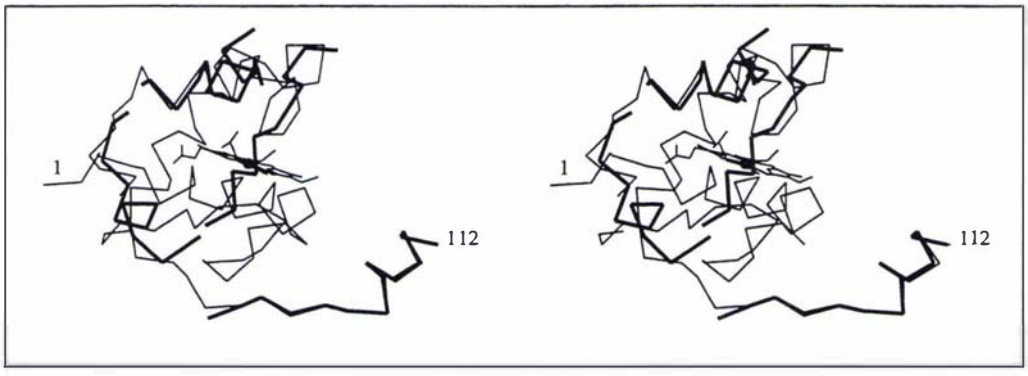



**Figure 6.10** Stereo view of a Cα superposition of the initial manually built N-terminal polyalanine model (dark lines) and the current frozen NIR N-terminal model (light lines).

The figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

### *6.6.5 Molecular replacement with the room temperature Photon Factory data*

Owing to the difficulties encountered in model building into the maps calculated from the frozen data sets the molecular replacement solution was recalculated in the room temperature (RT) cell with the Photon Factory data (NIR1). The search model used was the refined β-propeller dimer. The rotation function was computed with a Patterson integration radius of 30 Å, and data of resolution range 8.0-4.0 Å were used in both the AMORE rotation and translation function calculations. The results of the rotation and translation function calculations are given in Table 6.12 and Table 6.13.

**Table 6.12** Room temperature (NIR1) rotation function solutions

| Eulerian angles (°) | | | Peak Height |
|---|---|---|---|
| α | β | γ | (σ) |
| 301.6 | 90.0 | 189.2 | 47.3 |
| 301.6 | 90.0 | 9.1 | 47.1 |

Eulerian angles (α, β, γ) are defined such that α is a rotation of the coordinate system about z, β is a rotation about the new y axis, and γ is a rotation about the new z direction. The output angles are those required to orient the model after it has had its centre of mass placed at the origin and been aligned by its principle axes of inertia to the cell axes.

The two solutions for the dimer are clearly related by the NCS two-fold. No other peaks with heights of greater than 14σ were present in the rotation function.

**Table 6.13** Room temperature (NIR1) translation function solution

| Rotation Function Solution (Eulerian angles) (°) | | | Translation Function Solution (Fractional Coordinates) | | | Correlation Coefficient | R-factor |
|---|---|---|---|---|---|---|---|
| α | β | γ | x | y | z | | |
| 301.6 | 90.0 | 189.2 | 0.3452 | 0 | 0.3571 | 0.638 | 0.391 |

There were no other peaks of height greater than 20% of the value of this top peak. Rigid-body fitting gave a final solution with a correlation coefficient of 0.76 and an R-factor of 0.36 for data in the resolution range 8.0-4.0 Å. These calculations gave a position and orientation for the dimer that was slightly shifted, owing to the change in cell dimensions, when compared to the frozen cell (for a description of the crystal packing of the two cells see Section 7.7.1).

From SIGMAA-weighted (Read, 1986) electron density maps calculated with phases from this molecular replacement solution, density for one of the *c* heme groups could be unambiguously assigned. It was also possible to determine the position of the axial histidine ligand and the covalently-linked cysteine residues (Figure 6.11). The *c* heme and its strictly conserved binding motif from *P. aeruginosa* nitrite reductase were built into the density. When the entire N-terminal domain from the *P. aeruginosa* enzyme (with side-chains truncated at substitution sites, see Section 6.5.1) was superimposed on to the *c* heme-binding fragment, secondary structure elements could be placed into density (Figure 6.12) and the N-terminus of the β-propeller domain and the C-terminus of the *c* heme-binding domain were in close proximity, confirming that the *c* heme and its ligands were correctly oriented. Compared to the *P. aeruginosa* and *T. pantotropha* nitrite reductases the N-terminal

domain of *P. stutzeri* had undergone an approximately 80° rotation about the linker region relative to the β-propeller domain.

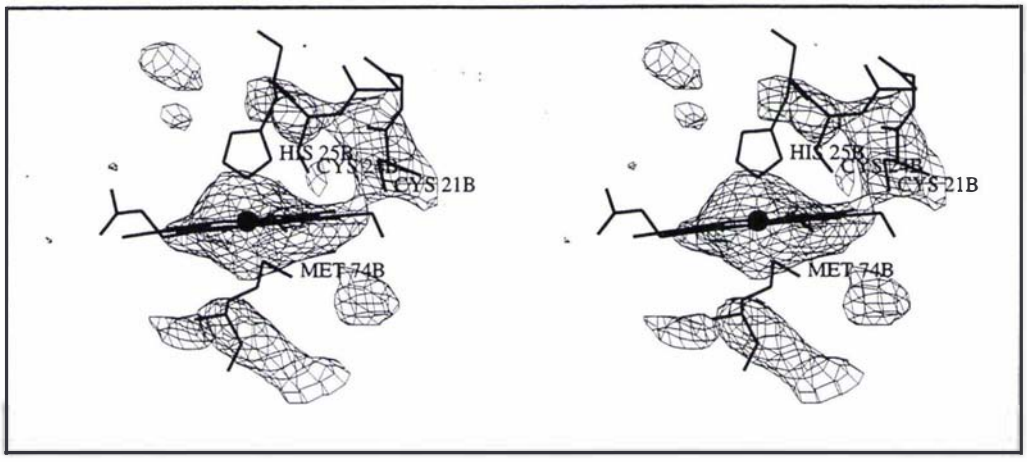



**Figure 6.11** Electron density for the *c* heme after molecular replacement in the room temperature data set.

The $2m|F_o| - D|F_c|$ map is contoured at 1σ, with phases calculated from the room temperature β-propeller dimer molecular replacement solution. The current RT model is overlaid. The figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

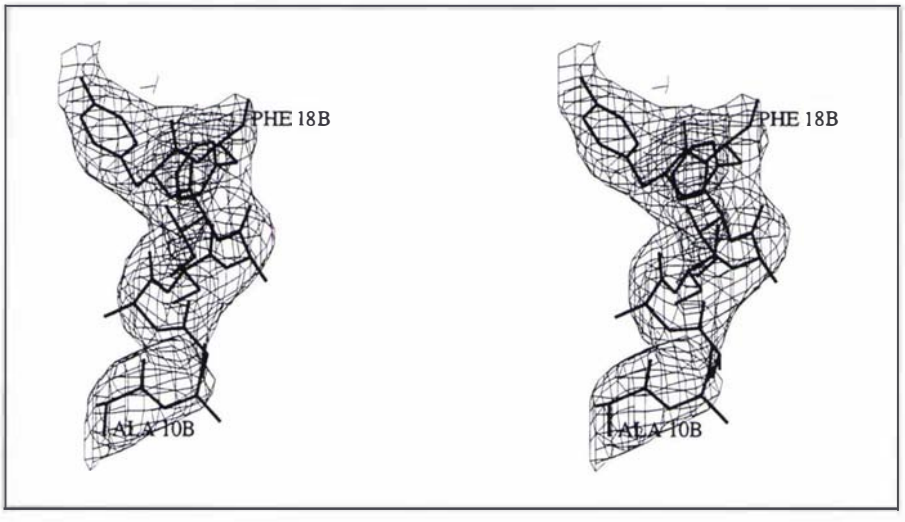



**Figure 6.12** Electron density for the N-terminal helix of the *c* heme-binding domain

The $2m|F_o| - D|F_c|$ map is contoured at 0.8σ, with phases calculated from the room temperature β-propeller dimer molecular replacement solution. The current model is overlaid. The figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

### *6.6.6 Locating the N-terminal domain of the frozen cell -Revisited*

The orientation and position of the N-terminal domain in the room temperature (RT) cell was used to help interpret the electron density maps for the higher-resolution frozen data set. The coordinates of the RT monomer B (N-terminal and C-terminal domains) were superimposed on to the monomer B frozen lattice β-propeller domain. The structure of the RT N-terminal domain was used as an aid in interpreting the frozen lattice electron density maps, phased from the β-propeller domains only. The partial, manually-built (Section 6.6.4), frozen N-terminal domain was excluded from the phase calculation to avoid biasing the maps. It transpired than an ~22° rotation was required to rotate the RT N-terminal coordinates into the electron density of the frozen lattice. Hence, based on the β-propeller domain, the N-terminal domain of monomer B had undergone a 22° rotation upon freezing.

This orientation was confirmed by ESSENS (Kleywegt and Jones, 1997), a program that detects structural features in electron density maps. In this algorithm the fragment to be found is rotated and translated within a specified region of the unit cell and the fit to an electron density map calculated. The molecular replacement *P. aeruginosa* N-terminal domain search model, (model (i) as described in Section 6.6.3), was used as the structural fragment for the search. The fit of this model was calculated to a region of SIGMAA-weighted $2m|F_o| - D|F_c|$ electron density map that had been masked (with a suitable safety margin) in the region of the room temperature N-terminal domain coordinates. The search was conducted on a grid of 1 Å spacing with the search model rotated in 10° steps. The results of this calculation yielded an unambiguous position and orientation for the N-terminal domain that was consistent with that found by visual map inspection. The solution found had a peak 10.8σ above the average correlation value (next highest peak 2.1σ above average). The orientation was further refined by searching around this highest peak in 2° steps, yielding a final solution of peak height 7.2σ (next highest 1.0σ). Maps calculated from the model including the newly positioned N-terminal domain showed the electron density around this domain to be weak, but consistent with it being correctly positioned.

From inspection of the frozen-crystal lattice symmetry it was clear that the second N-terminal domain (for monomer A) could not be positioned in such a way that it was related by the same NCS operator defined by the two β-propeller domains of the dimer. Rotating the newly positioned monomer B N-terminal domain by this two-fold rotation axis resulted in its NCS equivalent being involved in steric clashes with a crystallographically related symmetry equivalent β-propeller domain. The second N-

terminal domain (from monomer A) was found by (i) rotating about the β-propeller domain NCS two-fold into this sterically disallowed position, and (ii) then rotating further until the heme iron was positioned in the largest peak (in excess of 4σ) in the difference density ($m|F_o| - D|F_c|$) map. During this second step the C-terminus of the *c* heme-binding domain was kept close to the N-terminus of the β-propeller domain. This gave a solution for the second N-terminal domain (A) that packed well in the lattice, and was consistent with the features of both the $m|F_o| - D|F_c|$ and $2m|F_o| - D|F_c|$ maps phased from the other three domains of the asymmetric unit. ESSENS was also used to search for the position of this domain in a similar manner to that described previously for the first N-terminal domain (B). The top peak was consistent with the orientation found by manual rigid-body rotation, though the discrimination of this peak from the noise level was not so great as that of the first N-terminal domain (B). The top peak was 5.1σ above the average value with the next highest peak 1.5σ above the average. This orientation represented a 36° rotation from the orientation of N-terminal domain (B), relative to the β-propeller domain. Over the course of the refinement, density for many of the side-chains have appeared and these have been modelled. Omit maps have also been calculated at various stages of the rebuilding and refinement process to check that the subunit has been correctly placed. Though the density is weak and the temperature factors are high, the procedures listed above indicate that the second N-terminal domain (A) is correctly positioned. The free R-factor was monitored as the fragments of the N-terminal domains were included and refined. Cross-validation statistics for the inclusion of the two N-terminal domains in the frozen model, and the single domain in the room temperature model are listed below (Table 6.14).

**Table 6.14** Validation statistics for inclusion of the N-terminal domains for the frozen and room temperature NIR models

| Model | N-terminal domain | Free R | R |
|---|---|---|---|
| Frozen | A and B | 0.313 | 0.279 |
| | A | 0.319 | 0.286 |
| | B | 0.322 | 0.294 |
| | Neither | 0.323 | 0.300 |
| Room temperature | A | 0.295 | 0.239 |
| | None | 0.308 | 0.270 |

The free R-factor does decrease on inclusion of the N-terminal domains as described above, though in some instances the decrease is not very much. It is inter-

esting to note that the difference between the free R-factor and the conventional R-factor is slightly greater if the N-terminal domains are present, though this may reflect the fact that these statistics were calculated from models that had been refined with the N-terminal domains included.

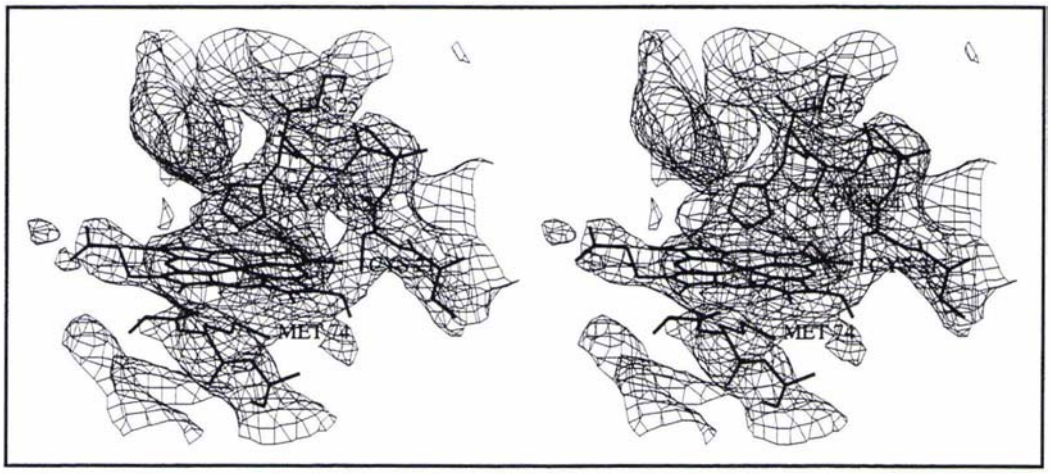



**Figure 6.13** Stereo view of the electron density for the NIR *c* heme pocket

The map is calculated from a recent set of refined room temperature NIR coordinates. The $2m|F_o| - D|F_c|$ map is contoured at 0.8σ. This figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

The second N-terminal domain (A) has not been modelled in the room temperature cell. Density is present in the area in which the corresponding domain is found in the frozen cell, but no large peak for the iron atom position is apparent. The best course is probably that this domain should be built by building smaller fragments, e.g. secondary structure elements and amino acid residues, rather than placing the major part of the whole domain, in order to confirm that the orientation found for the corresponding domain in the frozen lattice is indeed correct.

## 6.7 REFINEMENT OF THE NITRITE REDUCTASE STRUCTURE

### *6.7.1 Overview of the nitrite reductase refinement*

Refinement of the partial nitrite reductase model was initiated once the β-propeller domains had been placed in the cell by molecular replacement. Refinement and building of this model was conducted at the same time as the various procedures used to locate the N-terminal domains that have been described previously. It was hoped that improvements in the phases, resulting from corrections to the partial model from refinement and model building, would provide more information in the electron density maps and aid in the location of the missing parts of the structure. Refinement of a partial structure can be problematic as it cannot be determined how much of each observed structure factor amplitude is due to the missing part of the model. Maximum likelihood refinement provides a method of estimating the errors in the calculated structure factors owing to missing and misplaced atoms (see Section 3.5.7.4). Prior to and during the refinement of this structure, maximum likelihood refinement was implemented in some of the macromolecular refinement packages, and where possible these were used in the refinement.

The limited number of reflections, resulting from the moderate resolution of the diffraction data, meant that care had to be taken to avoid overfitting the model during the refinement. At the limited resolution of this analysis the ratio of observations to parameters is close to one. Although the R-factor can be lowered simply by adding more parameters this can result in no real improvement in the model. The free R-factor (Brunger, 1992a) was therefore used as a monitor to avoid overfitting the data. Non-crystallographic symmetry constraints or restraints between the two molecules in the asymmetric unit were also retained throughout the refinement, as an effective way of reducing the number of parameters.

The predominant problem in the nitrite reductase refinement, however, was the anisotropy of the data. It is important to apply some sort of correction for systematic effects such as this, that arise as a function of the crystal lattice rather than from the structures of the individual macromolecules contained within that lattice. If the anisotropy of the data is not taken into account then the refinement algorithm attempts to adjust the refined parameters, i.e. the atomic positions and temperature factors, to model this deviation in the measured intensities.

Unless otherwise specified all refinement was conducted using the frozen data set, initially from a single crystal (NIRf3) and later the combined data from two frozen

crystals (NIRc). During refinement, the models were restrained to the Engh and Huber stereochemical library (Engh and Huber, 1991).

### 6.7.2 Cross-validation

The free R-factor was monitored during refinement. Approximately 5% (1498 reflections) of the NIRf3 data were randomly selected for the free-R set. These free reflections were retained throughout all refinement, even when the combined data set (NIRc) was used. The same reflections were included in the free-R set of the almost isomorphous Photon Factory data set; here the total free-R set was also approximately 5% of the total data (969 reflections).

### 6.7.3 Map calculations and model building

After rigid-body refinement, maps with SIGMAA-weighted Fourier coefficients, ($2m|F_o| - D|F_c|$) and ($m|F_o| - D|F_c|$), (where $m$ is the figure of merit, and $D$ is an estimate of the error in the structure from coordinate errors) were computed (Read, 1986), and electron density maps were calculated using phases obtained from the model. All reflections were included in the map calculations. At several positions where the *T. pantotropha* and *P. stutzeri* nitrite reductase sequences differed, and hence the side-chain had been truncated to alanine, there were peaks in the difference density map indicating the position of the omitted side-chain. In the regions of the β-propeller domain where there were insertions in the *P. stutzeri* sequence relative to the search model density could be seen that indicated the trace of the main-chain. The presence of these peaks further validated the molecular replacement solution. Little density could be observed for the N-terminal domains however. After subsequent rounds of TNT and CNS maximum likelihood refinement, maps as described above were calculated, and the model was rebuilt using TURBO-FRODO (Roussel and Cambillau, 1991). A stereo view of the $2m|F_o| - D|F_c|$ electron density map calculated from a recent frozen NIR coordinate set is presented in Figure 6.14.

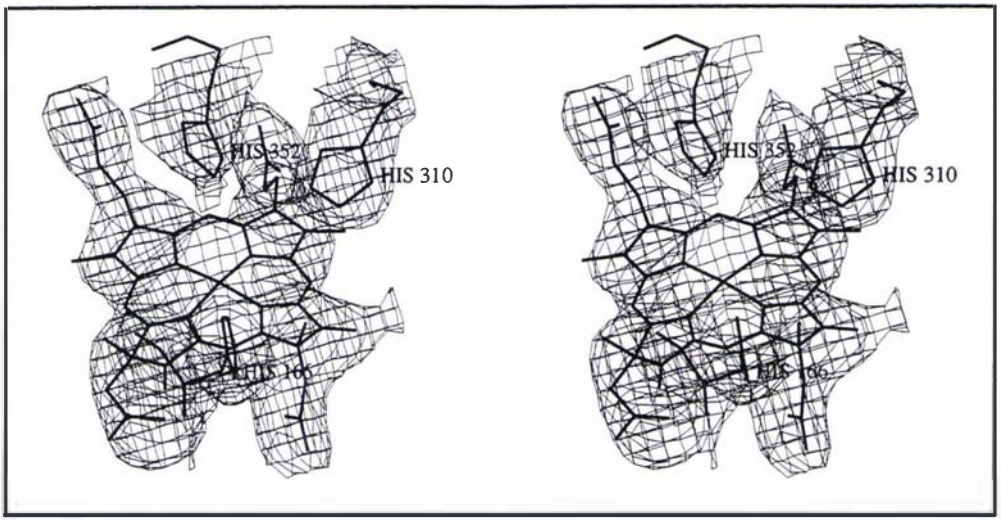


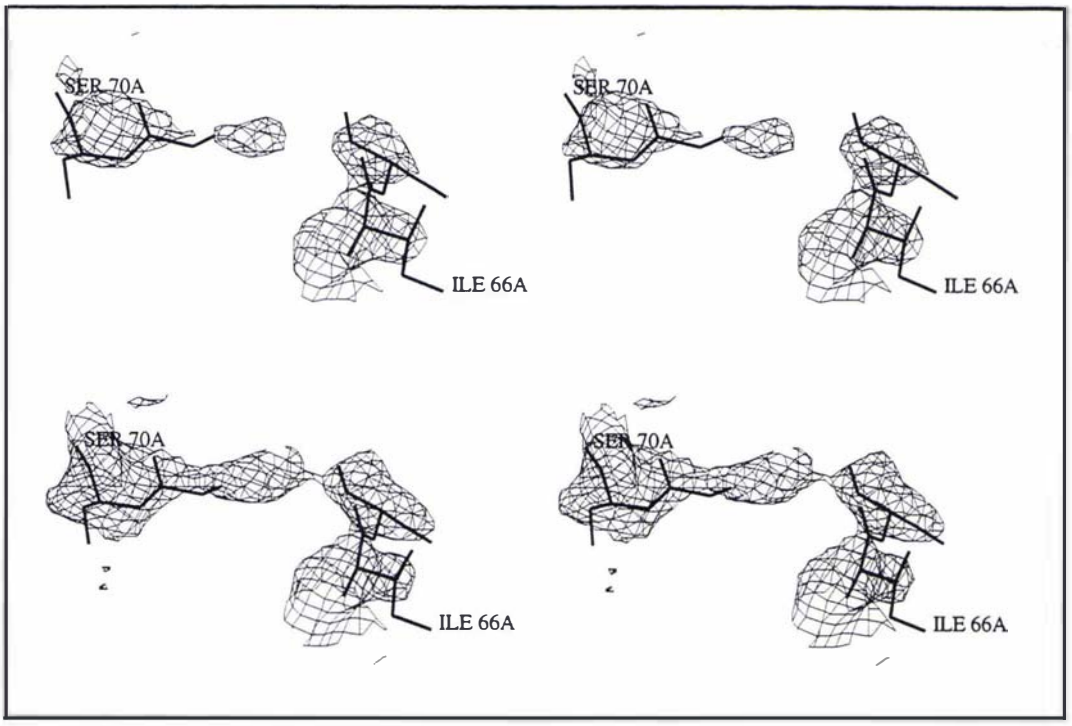

**Figure 6.14** Stereo view of the $2m|F_o| - D|F_c|$ map of the $d_1$ heme pocket

The $2m|F_o| - D|F_c|$ map is contoured at 1σ. This figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

Tests of maps calculated with different coefficients were conducted with REFMAC. The SIGMAA parameters *m* and *D* are calculated from the free set of reflections and hence are subject to less bias from the model. Any unobserved reflections in the native data set are given the value $D|F_c|$, and hence the $2m|F_o| - D|F_c|$ term becomes $D|F_c|$, and the contribution to the difference coefficients $m|F_o| - D|F_c|$ is zero. Tronrud (Tronrud, 1996b) and Cowtan (Cowtan, 1996) have shown that absent reflections can cause unpredictable noise in map calculations. Test calculations have shown that the restoration of these missing terms reduces systematic noise in the $2m|F_o| - D|F_c|$ maps (Murshudov et al., 1997). On the other hand, the use of these coefficients may introduce bias owing to the missing observations being replaced by weighted structure factors calculated from the model. Model bias in the electron density maps, by calculation of phases from the atomic model, is a major problem in the refinement of protein models determined by molecular replacement methods. Considering the relative weakness and incompleteness of the outer resolution shells of the native data set, the use of calculated structure factors in place of these missing reflections would have a relatively large impact on the $2m|F_o| - D|F_c|$ synthesis, and hence these calculated terms were not utilised further.

### 6.7.4 NCS averaging

At various intervals during the endeavours to locate the N-terminal domain NCS averaging methods (for a brief summary see Kleywegt, 1997) were used in attempts to improve the frozen-data electron density maps. Failure of this procedure was due to the difficulty of calculating a satisfactory molecular envelope for NCS averaging either by automatic (using DM, Cowtan and Main, 1996) or manual methods. It was also not possible to define the NCS operator for the missing domains. However, with the recent positioning of the second N-terminal domain it is has been possible to calculate molecular envelopes for NCS averaging procedures from the model coordinates. The NCS operators were calculated by superposition of the coordinates of the two N-terminal domains, and the C-terminal domains respectively. Because the dimer has different two-fold NCS operators for the N-terminal and C-terminal domains, a separate mask was calculated for each of these domains from their atomic model, using the algorithm MAMA (Kleywegt and Jones, 1994). A single round of NCS averaging without phase recombination was then conducted with CNS (Brunger et al., 1997). After averaging a $2m|F_o| - D|F_c|$ electron density map within the calculated NCS masks, the correlation coefficients were 0.72 and 0.99 for the N-terminal and C-terminal domains respectively. This protocol does improve the quality of the maps, particularly in the regions of the N-terminal domains (Figure 6.15).



Previous page.

**Figure 6.15** Comparison of nitrite reductase averaged and unaveraged $2m|F_o| - D|F_c|$ maps

The unaveraged map is shown above, and the averaged map below. Both maps were calculated from the same atomic model, and a region of the N-terminal domain is shown. This figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

### *6.7.5 Anisotropic correction*

To obtain a good scaling function for the nitrite reductase refinement it was important that a correction be applied for the strong anisotropy present in the diffraction data. If this is not done then the refinement protocols attempt to incorporate these systematic deviations in the intensities into the atomic model. For small molecules, where the observation/parameter ratio is large, anisotropic temperature factors can be refined. Because of the restricted observation to parameter ratio generally found for macromolecular refinement, however, an overall anisotropic temperature factor averaged over the unit cell can serve a similar purpose. Protocols that have been used to do this to date have been based on the use of Wilson statistics (Sheriff and Hendrickson, 1987, Blessing et al., 1996), or an analysis of the Patterson origin peak (Blessing and Langs, 1988). This temperature factor can be used to apply a correction to the native structure factors and then allow the refinement of an isotropic atomic model. The expression for the overall anisotropic temperature factor is:

$$B_{ani} = \frac{1}{4}\left( b_{11}a^{*2}h^2 + b_{22}b^{*2}k^2 + b_{33}c^{*2}l^2 + 2b_{12}a^*b^*hk + 2b_{13}a^*c^*hl + 2b_{23}b^*c^*kl \right)$$

*Eq. 6.3*

As the anisotropic temperature factor is averaged over the cell, then the parameters fitting this expression are restricted by the lattice symmetry. For the monoclinic nitrite reductase crystals, with the *b* axis-unique, $b_{12}=b_{23}=0$. The least squares fit of these parameters is achieved by correcting $|F_c|$ (derived from the isotropic model), by minimizing the difference with $|F_{obs}|$. Normally in scaling $|F_{obs}|$ to $|F_c|$, $1/k$ (where $k$ is the scale factor) is applied to $|F_{obs}|$ and the isotropic B-factor is applied to $|F_c|$. However for reasons of practicality the anisotropic correction is often applied to the observed reflections. An anisotropic correction has been implemented into the scaling function of several macromolecular refinement programs: REFMAC (Murshudov

et al., 1997), X-PLOR/CNS (Brunger, 1992b and Brunger et al., 1997), and SHELX-97 (Sheldrick, 1997).

To illustrate the importance of correcting for the anisotropy present in the nitrite reductase data sets, parallel structure factor calculations were conducted on recent frozen and room temperature models (Table 6.15).

**Table 6.15** Comparison of scaling functions for NIR models

| NIR model | Scaling model | Free R | R |
|---|---|---|---|
| Frozen | Isotropic | 0.394 | 0.356 |
| | Anisotropic | 0.314 | 0.279 |
| Room Temperature | Isotropic | 0.352 | 0.301 |
| | Anisotropic | 0.295 | 0.239 |

The calculation of the anisotropic correction is not trivial at moderate resolution, where Wilson statistics are not applicable, and with a model that is incomplete and likely to contain errors. The scaling protocol implemented in X-PLOR and CNS refines the anisotropic parameters and an overall scale factor by a least squares minimization followed by calculation of the bulk solvent correction and then further cycles of refinement of the scale factor and anisotropic correction.

Comparisons of the anisotropic parameters calculated for both the room temperature Photon Factory data and frozen data (Table 6.16) shows that the systematic differences in intensities relative to the cell axes are remarkably similar, with $a^*$ exhibiting the weakest diffraction, and $c^*$ the most intense.

**Table 6.16** Comparison of the overall anisotropic correction parameters for the frozen and room temperature data sets

| | $b_{11}$ | $b_{22}$ | $b_{33}$ | $b_{12}$ | $b_{13}$ | $b_{23}$ |
|---|---|---|---|---|---|---|
| frozen data | -30.0 | -5.2 | 35.2 | 0 | 4.1 | 0 |
| room temperature data | -34.4 | -0.54 | 34.9 | 0 | 7.1 | 0 |

It is curious to note that though the parameters for the anisotropic correction are slightly larger for the room temperature data (Table 6.16), application of anisotropic scaling improves the frozen model to a larger extent (Table 6.15). The anisotropic parameters calculated above are consistent with plots of $ln <F_{obs}>$ against resolution for data of constant *h*, *k* and *l* for both data sets (Figure 6.3 and Figure 6.8). If

similar plots are calculated with the observed amplitudes after correction for anisotropy it is clear that although the calculated parameters do account for part of the systematic deviation in the data, there are still marked relative differences in the intensities with respect to orientation, particularly with the higher resolution reflections. This may indicate that it is not possible to fully describe the anisotropy by a cell-averaged ellipsoid.

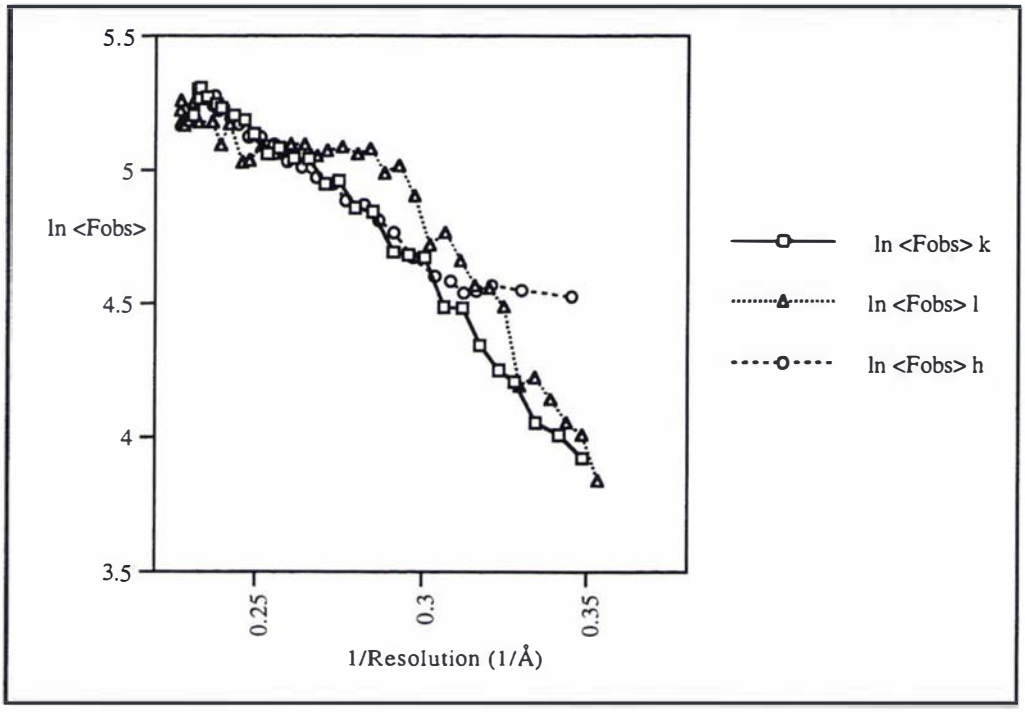



**Figure 6.16** Plot of the logarithm of the mean corrected amplitude of the frozen NIR data set as a function of resolution for reflections of constant *h*, *k* and *l*

This figure was prepared with DATAMAN (Kleywegt and Jones, 1996).

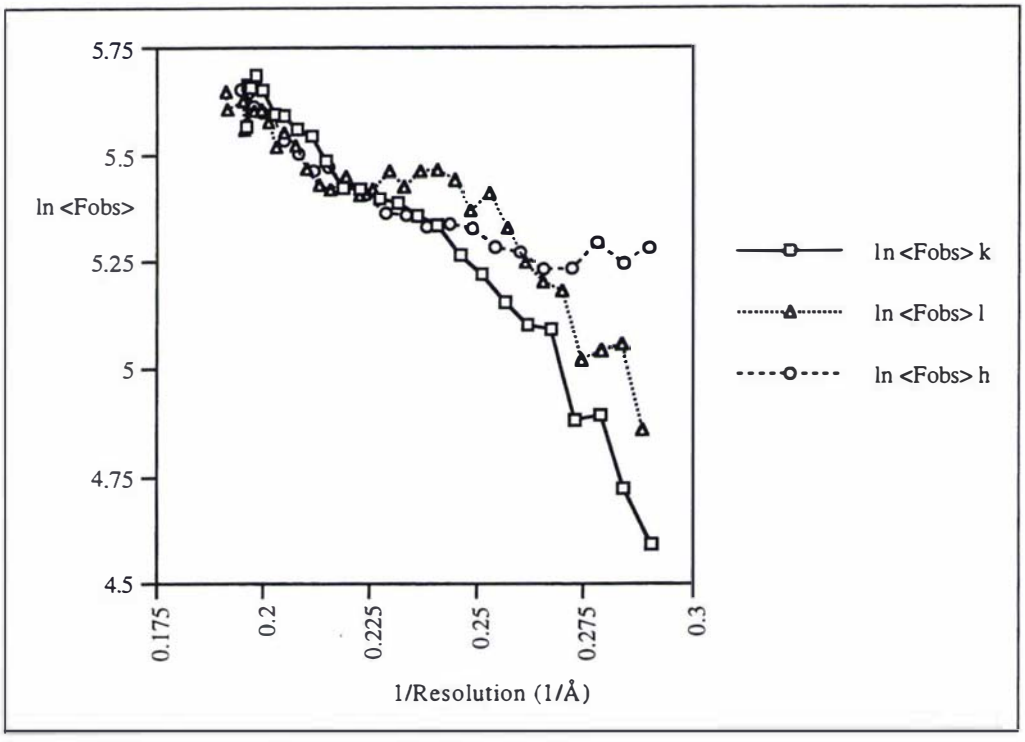



**Figure 6.17** Plot of the logarithm of the mean of the corrected amplitude of the room temperature data as a function of resolution for reflections of constant *h*, *k* and *l*

This figure was prepared with DATAMAN (Kleywegt and Jones, 1996).

### *6.7.6 Solvent model and bulk solvent correction*

The bulk solvent models for the treatment of the low resolution reflections have been discussed in Section 3.5.5. During the initial TNT refinement a Babinet-type solvent correction was applied. However, for the refinement conducted using X-PLOR and CNS the mask-type solvent correction was utilized. The bulk solvent correction was utilised in the refinement and map calculations for both the frozen and room temperature cells. It was necessary to manually set the value for the bulk solvent temperature factor ($B_{sol}$) for the scaling of the room temperature data.

During the latter stages of the refinement, ordered water molecules became visible in the electron density maps, predominantly in the region of the $d_1$ heme-binding domain. Ninety-nine solvent atoms have been included in the atomic model of the frozen NIR lattice. A water molecule was included in the model if a 3σ peak was present in the $m|F_o| - D|F_c|$ map, and the solvent atom formed a chemically sensible

hydrogen bond with the polypeptide. Upon refinement, water molecules were removed from the model if their temperature factors refined to high values (>80 $Å^2$), or the electron density did not remain in the $2m|F_o| - D|F_c|$ map once the water atoms were included in the phase calculation. The free R-factor was monitored to validate the addition of these solvent molecules (Table 6.17).

**Table 6.17** Validation of the solvent model

| Solvent Molecules | Free-R | R |
|---|---|---|
| 99 | 0.304 | 0.266 |
| None | 0.312 | 0.274 |

Statistics calculated on the current model

### *6.7.7 Temperature factor models*

During the early stages of refinement with the frozen data, with just the two β-propeller domains in the partial model, an overall temperature factor was refined. Once the two N-terminal domains had been positioned one temperature factor per residue was refined. The free R-factor was monitored to validate the inclusion of these additional parameters in the refinement. The effect on the free R-factor of refining a more sophisticated temperature factor model and hence a larger number of parameters was investigated. The modelling of two temperature factors per residue (one for side-chain atoms and one for main-chain atoms), or restrained individual isotropic temperature factors was found not to be appropriate for the current frozen lattice model (Table 6.18). An isotropic temperature factor was refined for the solvent atoms.

**Table 6.18** Temperature factor models for frozen NIR

| Temperature factor model | Free R | R |
|---|---|---|
| Grouped - 1 B per residue | 0.313 | 0.275 |
| Grouped - 2 B per residue | 0.315 | 0.272 |
| Individual isotropic | 0.314 | 0.276 |

Statistics calculated on a recent coordinate set.

Surprisingly, a grouped temperature factor model of two temperature factors per residue (one for side-chain atoms and one for main-chain atoms), was justified for the

room temperature refinement (Table 6.19), even though the data extend to a much lower resolution than that of the frozen lattice (3.4 Å and 2.8 Å respectively).

**Table 6.19** Temperature factor models for the room temperature NIR

| Temperature factor model | Free R | R |
|---|---|---|
| Grouped - 1 B per residue | 0.323 | 0.275 |
| Grouped - 2 B per residue | 0.319 | 0.286 |

Statistics calculated on a recent coordinate set.

### *6.7.8 Treatment of NCS*

In the early stages of the refinement, with the partial model containing just the two β -propeller domains, NCS constraints were enforced during the TNT maximum likelihood refinement, in order to minimize the number of parameters refined and to speed up the process of rebuilding these two large domains. When X-PLOR and latterly CNS were used to refine the structure, the NCS constraints were released, and tight NCS restraints enforced. This was because X-PLOR is limited in its application of NCS constraints in that only one NCS operation can be defined for a monomer. During these refinement steps the N-terminal domain was located, initially only for one monomer (B) but then eventually for the second (A). For nitrite reductase the NCS two-fold axis relating the two N-terminal domains was not coincident with that relating the β-propeller domains; thus the complete dimer could not be described in terms of a single NCS operation.

### *6.7.9 Progress of the frozen lattice nitrite reductase refinement*

#### *6.7.9.1 TNT rigid-body refinement*

The rigid-body fitting procedures in AMORE had been utilised to refine the positions of the two β-propeller domains, following molecular replacement. Further cycles of rigid-body refinement were carried out in TNT (Tronrud et al., 1987). The monomers were initially defined as rigid-body units, and were then broken up into fragments determined by the locations of insertions and deletions in the sequences of the two proteins. This refinement protocol resulted in a very small improvement in the model. The initial model had a free R-value of 0.458 and an R-factor of 0.435 for data in the range 20-2.8 Å. These values dropped slightly to 0.455 and 0.431 respectively.

*6.7.9.2 TNT maximum likelihood refinement*

The β-propeller domains were next subjected to maximum likelihood refinement in TNT (Tronrud et al., 1987, Pannu and Read, 1996). The Babinet solvent correction was applied and the model was refined with NCS constraints, and a single overall temperature factor. Several rounds of refinement and rebuilding, however, yielded little improvement of the model, with the free R-factor converging at 0.45 and the conventional R-factor at 0.39. It became clear that some sort of correction was necessary for the anisotropy of the data for refinement to be successful (see Section 6.7.5). Compounding this problem were the sequence errors identified later on in the refinement (Section 6.7.12); because the amino acid sequences indicated that these regions contained the greatest differences from the search model, rebuilding was concentrated on these areas, until it was realized that these sequence changes may be artefactual.

TNT contains an algorithm that outputs cell-averaged anisotropic parameters, but these are not applied during the refinement. X-PLOR was utilised to calculate an anisotropic correction for the frozen nitrite reductase data set and apply it to the observed structure factor amplitudes. This set of corrected observed data was then used for TNT maximum likelihood refinement. The anisotropic correction in conjunction with improvements in the model through model building and identification of sequence errors, resulted in the free R-factor decreasing to 0.40 and the R-factor to 0.34. At the same time as this partial model was being refined, attempts were made to locate the two missing N-terminal domains. The incorporation of the core of these two domains in addition to further iterative cycles of model building and refinement improved the model further (R-factor= 0.34 R-free= 0.38).

*6.7.9.3 CNS maximum likelihood refinement.*

With the recent implementation of a maximum likelihood target in the software package CNS, these algorithms were used in the later stages of the nitrite reductase refinement. The use of these programs allowed for the more simple calculation of the anisotropic correction as all the routines required were contained within the one software package. They also allowed for the implementation of the mask-calculated bulk solvent correction that had proved to be successful with the embryonic hemoglobin refinement (Section 3.5.5).

Refinement proved difficult, with often little improvement in the model (as monitored by the free R-factor) during cycles of refinement. Often it was not possible to refine

the atomic positions of the model after a session of manual rebuilding unless the temperature factors (which were either set to a uniform value of 50 $Å^2$ or were truncated to be less than 80 $Å^2$) were refined first. Refinement in this manner typically resulted in the temperature factors reaching high values for the N-terminal domain. It is not possible to determine whether the high temperature factors for this region are a result of coordinate error or due to high atomic motion in this region. Refining the temperature factors before the atomic positions (it is necessary to refine the atomic positions and temperature factors separately in CNS) is undesirable as it is not possible to position the atoms precisely during manual rebuilding and hence the temperature factors could act as an error sink for differences between the observed and calculated structure factors. However, in order to obtain a good scaling model for a successful refinement it did appear that the N-terminal atoms required large temperature factors. This behaviour was apparent with both the frozen single crystal (NIRf3) and combined (NIRc) data sets.

In order to determine whether the effective resolution of the frozen data was less than 2.8 Å owing to the weakness and incompleteness of the data in the outermost resolution shells, cycles of refinement and calculation of electron density maps were conducted using data to 3.4 Å, the resolution at which the refinement of the room temperature data appeared to be more successful. However refinement with the higher resolution terms omitted did not proceed any more smoothly, and there was no noticeable improvement in the quality of the electron density maps.

The final free R-factor and conventional R-factor are 0.304 (1498 reflections), and 0.266 (27724 reflections) respectively for all data in the range 40.0-2.8 Å. The model is tightly restrained, with rms deviations from ideal bond lengths of 0.0095 Å and ideal bond angles of 1.76°. Further refinement statistics are provided in Section 6.7.14.1.

### *6.7.10 Refinement using the room temperature data.*

Refinement at 3.4 Å is extremely problematic owing to the low number of observed reflections and the lack of detail in the electron density maps, though examples exist in the literature of successful refinements at this resolution by following careful protocols (Kleywegt et al., 1997). The N-terminal domain (B), first found for the room temperature data set, appears to be the most ordered of the three N-terminal domains (compared with the two from the frozen lattice). During the latter stages of

the refinement in the frozen cell the coordinates for the β-propeller domains and the N-terminal domain of monomer B of the frozen nitrite reductase model were superimposed on to the model of the room temperature cell and maximum likelihood CNS refinement was conducted. Anisotropic scaling and a bulk solvent correction were implemented during refinement, with a grouped temperature factor model refined (one temperature factor refined for the main-chain atoms and another for the side-chain atoms). Tight NCS restraints were imposed during refinement. This model currently has a free R-factor of 0.286 and an R-factor of 0.234 for all data in the range 40-3.4 Å.

Electron density maps calculated from this model with the room temperature data set, though naturally of a lower resolution, were of a higher quality showing new features in the maps for the N-terminal domain included in the model (monomer B), and also features for side-chains that were not included in the β-propeller model. These side-chains were added to the model and several loop regions rebuilt. Relatively strong connected electron density peaks in the map were present in the region where the N-terminal domain of monomer A was expected to be found, though they were not consistent with the exact trace of the domain if monomer A from the frozen lattice was superimposed on that of the room temperature cell (the superposition was calculated on the β-propeller domain for this monomer). No peaks consistent with the shape of the *c* heme group were readily apparent in the maps.

### *6.7.11 Cartesian and torsion angle molecular dynamics*

The use of molecular dynamics for the refinement of macromolecules (very briefly discussed in section 3.5.7.2 on page 62) has been successful in many cases (Brunger et al., 1987) and has resulted in the widespread use of such methods in structure refinement. A further development has been the introduction of simulated annealing procedures that refine the torsion angles of the model (Rice and Brunger, 1994). As only rotations about bonds are considered, the number of parameters refined is substantially reduced, making it a method more appropriate for refinement at moderate resolution. Torsion angle dynamics has been used successfully in the refinement of some difficult structures (Braig et al., 1995). Both atomic and torsion angle simulated annealing protocols were conducted on the frozen nitrite reductase model at various stages of the refinement described above. The R-factor would typically decrease after these procedures, but the free R-factor would increase indicating that the model was not improving. Similar experiences

have been reported using these protocols in the refinement of other proteins at moderate resolution (Kingston, 1996; S. Moore, personal communication).

Curiously, when torsion angle molecular dynamics protocols were conducted on the room temperature model the procedure was partially successful. The free R-factor would decrease (to a similar value reached with maximum likelihood positional refinement) indicating that the model was improving to a certain degree, but the difference between this statistic and the conventional R-factor would increase (compared to a standard maximum likelihood protocol) suggesting overfitting of the model to the data.

### *6.7.12 Sequence conflicts*

It was clear from early in the refinement that there appeared to be differences in the sequence of the protein, as indicated by the electron density maps, compared to the previously published sequence for the *P. stutzeri* JM300 nitrite reductase (Smith and Tiedje, 1992). Although these maps were generally not of sufficient quality to discriminate between different side-chains, it was clear that in some instances the deletions suggested by the sequence alignment (Figure 5.4, on page 101) were not present in the crystal structure. Initially, attempts were made to model the published sequence, but the coordinate shifts from refinement and the electron density maps were consistent with these deletions not being present. The sequence indicated by the electron density maps was compared with the highly similar nitrite reductase sequence from *P. stutzeri* ZoBell (Jungst et al., 1991). The major regions of sequence anomalies are given in tables below. Where possible, attempts have been made to rationalize these sequence differences in terms of possible frame shift errors in the nucleotide sequencing arising from base pair compressions. These sequencing artifacts are particularly common in G and C-rich regions (Gough and Murray, 1983). In the tables given below, the published *P. stutzeri* JM300 nitrite reductase sequence is given, followed by the proposed amino acid sequence indicated from the Fourier maps. For some of these regions the published nucleotide sequence is given, followed by a modified sequence (with insertions in bold) that could account for the sequence determined from the crystal structure. The sequences of the homologous enzymes from other species are also given. These regions all occur in the β-propeller domain, and although the current model is not complete in the N-terminal domain it indicates that the native protein contains 533 residues and not 525. All residue numbering now refers to the new sequence.

*6.7.12.1 Residues 119-127*

The region 119-127 (Table 6.20) is not as well conserved as some of the other regions where sequence anomalies exist and hence it is more difficult to rationalize the sequence changes observed. There clearly is no two-residue deletion in this segment of the polypeptide, but it is not so easy to determine the identity of some of the individual amino acids.

**Table 6.20** Sequence differences: residues 119-127

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *P. stutzeri* JM300 | E | R | R | --- | --- | R | Q | M | N |
| Proposed X-ray seq. | E | R | R | A | K | S | Q | M | N |
| *P. stutzeri* ZoBell | E | K | R | V | T | K | Q | M | N |
| *P. aeruginosa* | E | D | R | P | K | K | Q | L | N |
| *T. pantotropha* | E | D | R | P | T | Q | Q | E | N |
| *A. eutrophus* | A | Q | R | P | T | Q | K | M | N |

seq. is an abbreviation for sequence

Residue 123 is clearly a large side-chain and has been modelled as a lysine in the *P. stutzeri* structure. The identity of residue 124 is not clear; the homologous sequences have a lysine or glutamine at this position. Side-chain density is observed for this residue past the gamma position, when it is modelled as a serine, indicating that there is a residue with a large side-chain at this position, though it is not possible to identify the exact nature of the amino acid (Figure 6.18).

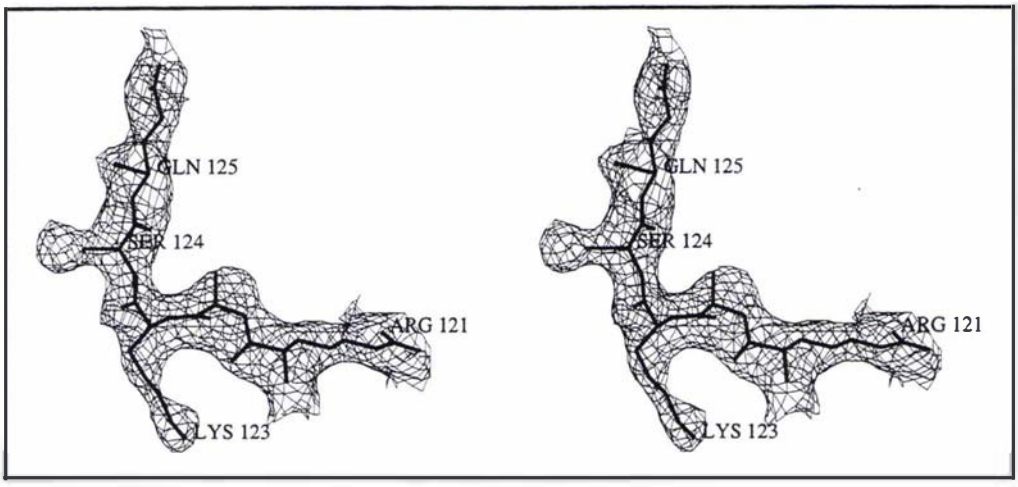



**Figure 6.18** Stereo view of the $2m|F_o| - D|F_c|$ map for residues 121-125

The density map is contoured at 1σ, and truncated 2.2 Å from the displayed atoms for clarity. This figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

*6.7.12.2 Residues 139-150*

Even from the earliest stages of refinement it was clear that the three-residue deletion proposed in the published sequence for the region now numbered 139 -150 (Table 6.21) was not consistent with the electron density maps (Figure 6.19). This part of the sequence corresponds to the region between the first and second β-strands of the β-propeller domain, an area that appears to be generally well conserved amongst the nitrite reductases of different species. Given the good quality of the electron density and the presence of easily recognisable residues (Gly, Lys) considerable confidence can be placed in this correction to the sequence.

**Table 6.21** Sequence differences: residues 139-150

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *P. stutzeri* JM300 | L | R | D | A | --- | --- | --- | Q | L | W | D | G |
| Proposed X-ray seq. | L | R | D | A | G | K | L | A | L | V | D | G |
| *P. stutzeri* ZoBell | L | R | D | A | G | K | L | A | L | I | D | G |
| *P. aeruginosa* | L | R | D | A | G | Q | I | A | L | V | D | G |
| *T. pantotropha* | L | R | D | A | G | Q | I | A | L | I | D | G |
| *A. eutrophus* | L | R | D | A | G | E | V | A | L | I | D | G |

seq. is an abbreviation for sequence

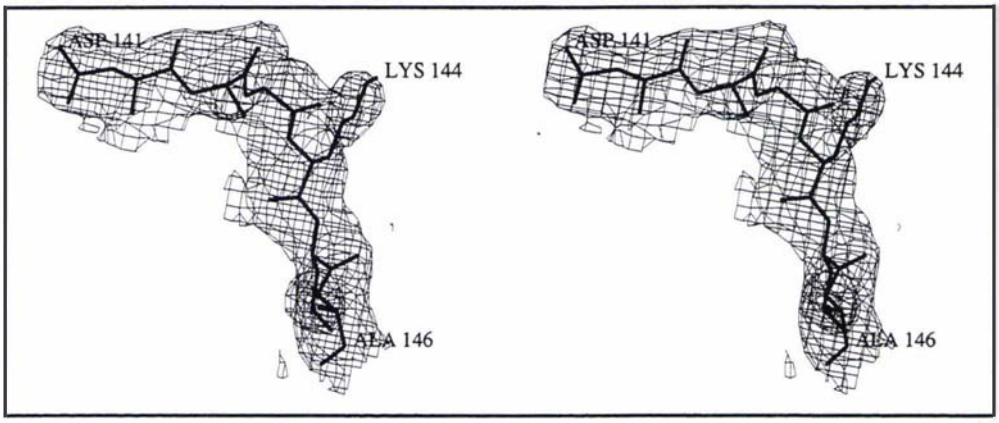



**Figure 6.19** Stereo view of the $2m|F_o| - D|F_c|$ map for residues 141-146

The density map is contoured at 1σ, and truncated 2.0 Å from the displayed atoms for clarity. This figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

### *6.7.12.3 Residues 174-187*

The region of the structure between residues 174-187 also exhibits a large degree of sequence similarity among different nitrite reductases and contains amino acids with distinguishable side-chains (Table 6.22). This region was initially modelled in terms of the published sequence, however problems were encountered in the region of the one-residue deletion (residue 183). Large positive difference density peaks were observed at the end of the side-chains of Thr 178, and originating from the Cα of Gly 182. Residue 178 is a tyrosine in the *P. stutzeri* ZoBell and *A. eutrophus* NIR sequences. The electron density for the side-chain at this position is consistent with this residue being tyrosine. The amino acid at position 182 is a conserved arginine residue in all the other nitrite reductase sequences, and in the *P. aeruginosa* and *T. pantotropha* structures forms an ionic interaction with a propionate of the $d_1$ heme. The difference density observed in the maps was consistent with this residue being arginine, forming this conserved interaction. Another structurally conserved interaction is formed by Asp 183, which participates in an Asx turn (described by Rees et al., 1983). The side-chain density of residue 176 when initially modelled as a methionine was flattened in one dimension, uncharacteristic of methionine, but consistent with this side-chain being tyrosine. This observation not only increased the confidence in the proposed frameshift errors, but also gave increased confidence that the electron density maps contained information from the observed intensities, and were not overly biased by the phases calculated from the model.

**Table 6.22** Frameshift errors: residues 174-187

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *P. stutzeri* JM300 | G | R | M | S | T | P | S | A | G | --- | W | L | T | T |
| Proposed X-ray seq. | G | R | Y | V | Y | T | V | G | R | D | G | L | T | T |
| *P. stutzeri* JM300 | g g c | c g t | a t g | t c t | a c a | c c g | t c g | g c c | g g a | - - - | t g g | c t g | a c c | a c c |
| Proposed nuc. seq. | g g c | c g **g** | t a t | g t c | t a c | a c c | g t c | g g c | c g g | **g** a t | g g **g** | c t g | a c c | a c c |
| *P. stutzeri* ZoBell | G | R | Y | V | Y | T | T | G | R | D | G | L | T | T |
| *P. aeruginosa* | G | R | Y | L | L | V | V | G | R | D | A | R | I | D |
| *T. pantotropha* | G | R | Y | L | F | V | V | G | R | D | G | K | V | N |
| *A. eutrophus* | G | R | Y | L | Y | V | V | G | R | D | A | R | L | D |

seq. is an abbreviation for sequence, nuc. is an abbreviation for nucleotide

A stereo view of the electron density present for the interaction of Arg 182 with the $d_1$ heme propionate is given in Figure 6.20.

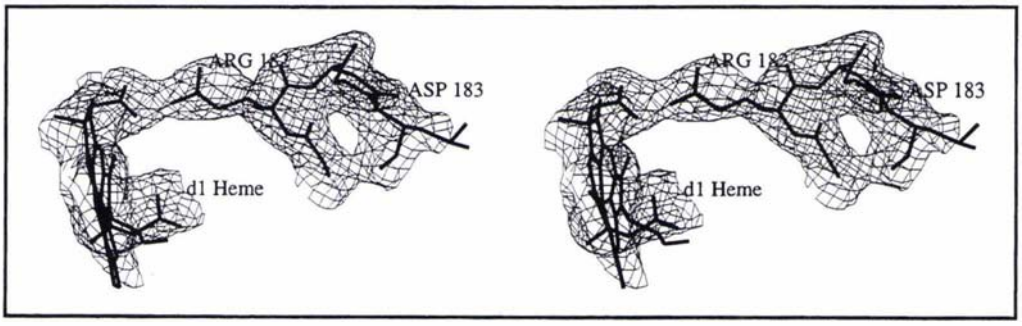



**Figure 6.20** Stereo view of Arg 182...$d_1$ heme propionate interaction

The $2m|F_o| - D|F_c|$ density map is contoured at 1σ, and truncated 2.0 Å from the displayed atoms for clarity. This figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

### *6.7.12.4 Residues 203-212*

A small difference occurs for the region 203-212 (Table 6.23). Again, density is apparent at the site of the proposed one-amino acid deletion. Nice density is also observed for residue 207. This is proline in the published sequence, but in the electron density it can be modelled much better as aspartic acid. It then makes a salt bridge with Arg 209, matching the Glu...Arg salt bridge found in the equivalent position for the *T. pantotropha* and *P. aeruginosa* enzymes (Figure 6.21).

**Table 6.23** Frameshift errors: residues 203-212

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *P. stutzeri* JM300 | R | L | G | --- | P | I | R | S | V | D |
| Proposed X-ray seq. | R | L | G | S | D | A | R | S | V | D |
| *P. stutzeri* JM300 | c g c | c t g | g g t | - - - | c c g | a t c | c g c | t c g | g t g | g a c |
| Proposed nuc. seq. | c g c | c t g | g g **g** | t c c | g a t | c ? ? | c g c | t c g | g t g | g a c |
| *P. stutzeri* Zobell | R | F | G | S | D | M | R | S | V | D |
| *P. aeruginosa* | K | I | G | I | E | A | R | S | V | E |
| *T. pantotropha* | K | I | G | S | E | A | R | S | V | E |
| *A. eutrophus* | R | V | G | M | E | A | R | S | D | G |

seq. is an abbreviation for sequence, nuc. is an abbreviation for nucleotide.

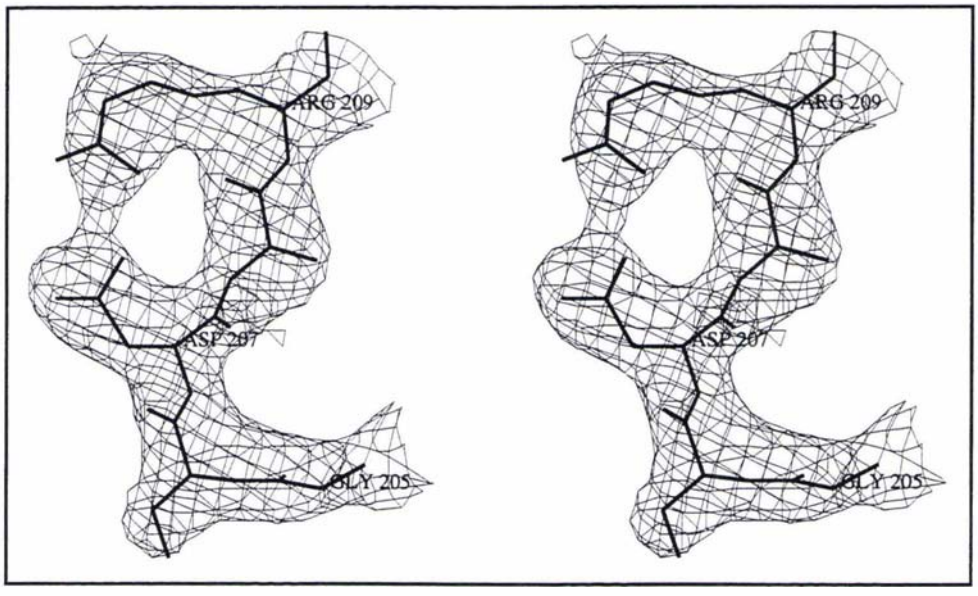



**Figure 6.21** Stereo view of the $2m|F_o| - D|F_c|$ map for residues 205-209

The density map is contoured at 1σ, and truncated 2.0 Å from the displayed atoms for clarity. This figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

### *6.7.12.5 Residues 349-360*

Residues 349-360 run across the top of the $d_1$ heme and hence are likely to be important in determining the environment of the heme (Table 6.24). The conserved histidine 352 has been implicated as being important for the reaction mechanism. The substitutions in this region suggested by the published sequence would have had far-reaching implications for the function of this protein. From the X-ray structure determination it does appear that this region is structurally conserved among the members of this protein family.

**Table 6.24** Frameshift errors: residues 349-360

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *P. stutzeri* JM300 | K | I | R | T | R | T | R | ---- | ---- | N | F | V |
| Proposed X-ray seq. | K | I | P | H | P | G | R | G | A | N | F | V |
| *P. stutzeri* JM300 | a a g | a t c | c g c | a c c | c g g | a c g | c g g | - - - | - - - | a a c | t t c | c t g |
| Proposed nuc. seq. | a a g | a t c | c **c** g | c a c | c c **c** | g g a | c g c | g g **g** | **g c a** | a a c | t t c | c t g |
| *P. stutzeri* ZoBell | K | I | P | H | P | G | R | G | A | N | F | V |
| *P. aeruginosa* | K | T | P | H | P | G | R | G | A | N | F | V |
| *T. pantotropha* | Q | T | P | H | P | G | R | G | A | N | F | V |
| *A. eutrophus* | K | T | P | H | P | G | R | G | A | N | F | M |

seq. is an abbreviation for sequence, nuc. is an abbreviation for nucleotide

A stereo view of the electron density present for this region of the protein is given in Figure 6.22.

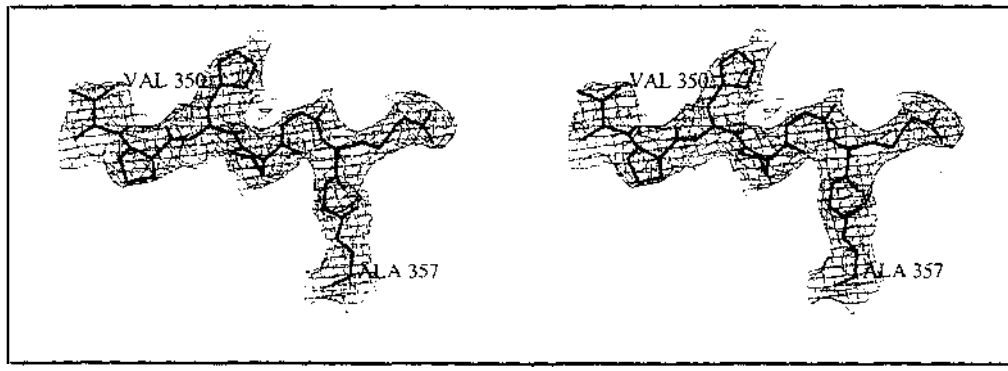



**Figure 6.22** Stereo view of the $2m|F_o| - D|F_c|$ map for residues 350-357

The density map is contoured at 1σ, and truncated 2.0 Å from the displayed atoms for clarity. This figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

### *6.7.12.6 Residues 488 and 332-333*

Another position where there appears to be a sequence anomaly is at residue 488, which is glycine in the published sequence, but a conserved tryptophan in the other five published nitrite reductase sequences. Examination of both $2m|F_o| - D|F_c|$ and $m|F_o| - D|F_c|$ electron maps reveal the presence of density for a large side-chain for the amino acid at this position, inconsistent with this residue being glycine. The surrounding residues fit the density well, indicating that the model is not out of register.

A strong argument for the conservation of this tryptophan can be made not only on the evidence above, but also as it forms a hydrogen bond to the carbonyl oxygen (OMC) of the $d_1$ heme C pyrrole.

It also appears that the insertion of a Pro between residues 332 and 333 is an artifact. No other nitrite reductase sequence contains this insertion and there is no room to accommodate this extra residue in the model.

It should be emphasized that the nucleotide frameshift changes illustrated in the tables above are only suggestions of what might have happened in the sequencing of this gene. Though the *P. stutzeri* JM300 and ZoBell nitrite reductase sequences appear to be more similar than previously thought, there will be most likely still be differences between them. Determining the sequence of a protein from electron density maps is impossible at 2.8 Å resolution; some possibilities can be discounted, e.g. small and large side-chain substitutions, but side-chains of similar size cannot be distinguished.

The molecular weight of the *P. stutzeri* nitrite reductase as determined by triple quadropole electrospray mass spectrometry is 59884 Da (±0.01%). This value differs with the figure of 54 kDa as determined by SDS-PAGE analysis (Weeg-Aerssens et al., 1991). The calculated molecular weight of the published *P. stutzeri* JM300 nitrite reductase sequence (including the hemes) is 59874 Da, a molecular weight more consistent with the experimentally determined Mw than the proposed X-ray sequence which has a calculated mass of 60491 Da (with both hemes included). However, it is unclear whether the $d_1$ heme is retained in the protein during the ionization of the electrospray analysis. If it is assumed that the $d_1$ heme is lost from the protein, then comparisons of the measured Mw with the X-ray sequence, minus the $d_1$ heme, reveal that the Mw for the X-ray sequence (59781 Da) is too low. For comparison, the calculated Mw for *P. stutzeri* ZoBell NIR (with both heme groups included) is 60859 Da. In order to determine the *P. stutzeri* JM300 nitrite reductase sequence unequivocally, and resolve these ambiguities, the gene needs to be resequenced.

I am very grateful to Dr. G. E. Norris and Ms J. Mudford for conducting the mass spectrometry analysis.

### 6.7.13 Summary

An unambiguous molecular replacement solution was obtained using the β-propeller domains of the dimer as a search model. However, location of the N-terminal domains was difficult, owing to the disorder exhibited by these regions of the molecule. The phasing power from the $d_1$ domains should have been sufficient to locate these domains if they were ordered in the lattice, with the missing part of the structure comprising less than 20% of the total scattering.

The structure of *P. aeruginosa* NIR was solved by initially placing the $d_1$ domains in the cell by molecular replacement. The N-terminal domains were visible in the electron density maps after rigid-body refinement of the $d_1$ domains, and a dimer of residues 30-543 could be built (with residues 80-100 and 450-470 omitted) before any individual atomic refinement was undertaken (Nurizzo et al., 1997). Clearly the phases calculated from the $d_1$ heme domains were accurate enough to allow the location of the *c* heme-binding domains in this case.

Partial models have been used for successful molecular replacement structure determinations that have comprised much less than the 80% that was present in this case, for example the structure of the human protective protein was solved by molecular replacement and two-fold averaging from an initial solution comprising 60% of the scattering (Rudenko et al., 1996); and the yeast cytochrome *c* peroxidase - cytochrome *c* complex was solved from a molecular replacement solution that placed just the peroxidase molecules (75% of the scattering matter) in the cell (Pelletier and Kraut, 1992).

Clearly it is not just a question of how much of the total scattering a model comprises that governs the quality of the electron density map. The similarity between the structures of the native molecule and the search model that is used to calculate the initial phases is also very important. In this regard the early maps for the *P. stutzeri* nitrite reductase β-propeller domains were actually very good, allowing the identification of several sequence errors, and the placement of side-chains that were not included in the search model.

Maps calculated from models which contained the N-terminal domains consistently exhibited very weak density for these regions of the molecule. This made rebuilding these regions a very difficult and laborious procedure. Omit maps were frequently calculated for fragments contained within the N-terminal region to confirm that they

were correctly positioned, and the free R-factor was monitored at all times during the cycles of model building and refinement. Refinement of the frozen model was problematic both before and after the N-terminal domains were modelled. The free R-factor would typically decrease only slightly during refinement after a session of rebuilding. The maps did, however, contain information that allowed the manual refitting of the model to the electron density. A major concern during the refinement was the appropriateness of the anisotropic scaling function. Some sort of correction was clearly necessary for the severe anisotropy present in the data, but it was unclear quite what effect the missing regions of the structure and coordinate errors would have on the effectiveness of such a correction. Including a model for the anisotropy did result in a dramatic drop in the free R-factor (Table 6.15) and improvements in the quality of the maps. The temperature factors for the N-terminal domains have refined to high values (see Table 6.26 and Table 6.27). The atomic temperature factor contains contributions both from atomic movements or disorder and from error in the coordinates. The low level of the electron density and the overlap of the density over several atoms as a result of high temperature factors makes positioning and refining the atomic coordinates very difficult (Ten Eyck, 1995). From inspection of the electron density maps for the current frozen model there are still ambiguous regions in both the N-terminal domains and coordinate error undoubtedly contributes to the high temperature factors. The recent success of NCS averaging procedures yields hope that the models for the N-terminal domains can be finished, though the completion of this refinement still appears difficult.

The weak density and high temperature factors have lead to the conclusion that the N-terminal domains may be disordered in the frozen crystal lattice. There certainly appears to be room within the arrangement of molecules in the crystal for this region of the protein to move, and disorder of protein subunits and/or domains is not uncommon (see Section 7.7.2). However it is certain that the lack of well-measured high resolution data (itself an effect that is probably due to disorder in parts of the molecule) contributed to the problems encountered during this structure determination. The problem of why the model would not refine as well against the frozen data as the room temperature data, and why there appeared to be more information in the unfrozen electron density maps remains unclear. Both data sets exhibited a similar anisotropy, and the success of the correction applied to the room temperature data would appear to validate this treatment as being appropriate. One major difference between the two data sets is the much greater mosaicity of the frozen crystals (1.1°), compared to that of the room temperature crystal (0.4°).

A return to the refinement of the room temperature model has been successful, with the refinement appearing to proceed more smoothly. The room temperature data are of lower resolution, but as well as being less mosaic these data were collected on a synchrotron source and may be more accurately measured. Density peaks visible in the maps for the N-terminal domain of monomer A are not consistent with the position of this domain found for the frozen model, suggesting that, like the N-terminal domains for monomer B, this domain is in a different orientation for the room temperature and frozen models.

In hindsight it is clear that the best approach to this problem would have been the collection of a better native data set, if this indeed was possible. The poor quality of the data made it extremely difficult to refine the model. The collection of well measured Friedel pairs could have yielded information on the position of the *c* heme iron atoms. It also would have helpful to have independent phase information and perhaps attempts to collect heavy atom derivatives should have been made a priority instead of persevering with what turned out to be a difficult molecular replacement problem. Phases from even a single poor low resolution derivative in combination with phases calculated from the β-propeller domains may have provided additional information in the maps.

If the difficulties of locating the frozen N-terminal domains are a result of disorder, this makes it difficult to model the structure with an atomic model and temperature factors. This type of model does not describe the overall dynamic behaviour of the molecule and it may be that a multiple conformer molecular dynamics refinement (Gros et al., 1990, Kuriyan et al., 1991, Wall et al., 1997) may prove to be necessary to model the structure of nitrite reductase in this frozen lattice form.

### *6.7.14 Quality of the current models*

#### *6.7.14.1 Frozen lattice*

The current model of the *P. stutzeri* nitrite reductase dimer in the frozen lattice comprises two complete β-propeller domains and two partially modelled N-terminal domains. The β-propeller domains of both monomer A and B comprise the complete polypeptide chain (residues 100-533) and the $d_1$ heme. The N-terminal domain for monomer B (residues 1-99) has residues 43-52 and 71 missing. Residues 43-52 comprise part of the large insertion in the N-terminal domain found only for the *P.*

*stutzeri* NIRs, and residue 71 is located in a loop just before the *c* heme ligand Met 74. The N-terminal domain for monomer A is less complete; residues 1-9, 33, 39-40, 44-56, 76-79 and 86-87 have not been included in the model (Table 6.25). The *c* hemes for both monomers have been modelled, though for the heme of monomer A the propionate side-chain of the A pyrrole ring has not been included. A number of amino acids, particularly in the N-terminal domains, have been modelled as alanine at this stage with no electron density visible for the side-chain atoms.

**Table 6.25** Atoms included in the model of NIR in the frozen lattice

| Atoms | Monomer A | Monomer B |
|---|---|---|
| polypeptide | 3746 | 3740 |
| *c* heme | 38 | 42 |
| $d_1$ heme | 48 | 48 |
| water | 99 | |

The free R-factor for the current frozen model is 0.304 (1 498 reflections), with a conventional crystallographic R-factor of 0.266 (27 724 reflections), for all data in the range 40.0-2.8 Å. The model is tightly restrained, with rms deviations from ideal bond lengths of 0.0095 Å and ideal bond angles of 1.76°.

The main-chain torsion angles ($\phi$,$\psi$) were not restrained during the refinement (though a Ramachandran plot was referred to during model rebuilding) and hence the distribution of these values provide some validation of the structure. A Ramachandran plot of main-chain torsion angles calculated with PROCHECK (Laskowski et al., 1993) shows that 96.7% of residues lie in the most favoured and additionally allowed regions, 3.0% are present in the generously allowed region, and 0.3% in the disallowed area (Figure 6.23). The three residues in the disallowed regions, 72 (monomer A), 196 (monomer B) and 462 (monomer B), are in loops with weak density.

The non-crystallographic symmetry was tightly restrained, with rms differences for the groups of NCS-restrained atoms less than 0.035 Å.

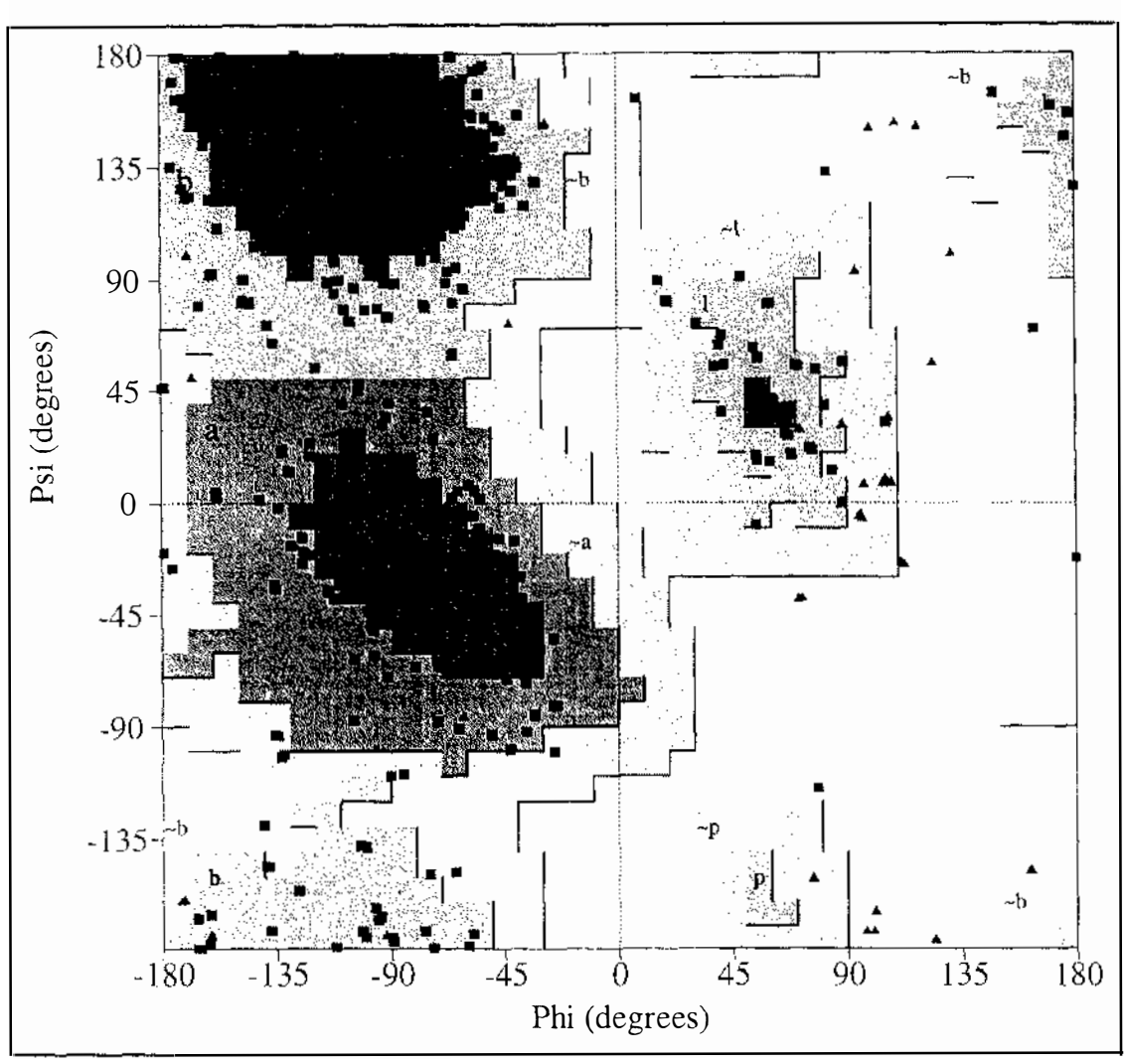



**Figure 6.23** Ramachandran plot for the frozen nitrite reductase dimer

Glycine residues are represented as triangles, with all other residues shown as squares. The plot was calculated with PROCHECK (Laskowski et al., 1993).

The average temperature factor the for the nitrite reductase dimer is 67 Å$^2$, comparable to the Wilson B value of 70 Å$^2$, though the use of Wilson statistics is not really justified with moderate-resolution anisotropic data. Temperature factors for the heme groups and protein atoms on a domain basis are given in Table 6.26.

**Table 6.26** Average temperature factors (Å$^2$) for the frozen nitrite reductase

| Atoms | Monomer A | | Monomer B | |
|---|---|---|---|---|
| | N-terminal domain | C-terminal domain | N-terminal domain | C-terminal domain |
| polypeptide | 110 | 51 | 95 | 72 |
| *c* heme | 125 | | 111 | |
| $d_1$ heme | 57 | | 84 | |

*6.7.14.2 Room temperature lattice*

The current model of the *P. stutzeri* nitrite reductase dimer in the room temperature lattice comprises two complete β-propeller domains and one partially modelled N-terminal domain. The β-propeller domains of both monomers comprise the complete polypeptide chain (residues 100-533) and their $d_1$ heme group. The N-terminal domain for monomer B (residues 2-99) has the majority of the fold modelled, along with the *c* heme group, with only the N-terminal residue 1, residues 43-52, and 72 missing (Table 6.27). Residues 43-52 comprise part of the large insertion in the N-terminal domain found only for the *P. stutzeri* NIR, and residue 72 is located in a loop just before the *c* heme ligand Met 74. A number of amino acids, particularly in the N-terminal domain have been modelled as alanine at this stage, with no electron density visible for the position of the side-chain atoms. For monomer A, five residues of the C-terminal region of the *c* heme-binding domain (residues 94-99), that are visible extending from the β-propeller domain, are included in the model.

**Table 6.27** Atoms included in the model of NIR for the room temperature lattice

| Atoms | Monomer A | Monomer B |
|---|---|---|
| polypeptide | 3351 | 3721 |
| *c* heme | -- | 38 |
| $d_1$ heme | 48 | 48 |

The free R-factor for the current room temperature model is 0.286 (969 reflections), with a conventional crystallographic R-factor of 0.234 (18 239 reflections), for all data in the range 40.0-3.4 Å. The model is tightly restrained, with rms deviations from ideal bond lengths of 0.010 Å and ideal bond angles of 1.79°.

The non-crystallographic symmetry was tightly restrained with rms differences for the groups of NCS-restrained atoms less than 0.012 Å.

As this model was derived from the molecule in the frozen lattice the distribution of the main-chain torsion angles ($\phi$,$\psi$) in a Ramachandran plot is very similar. A Ramachandran plot calculated with PROCHECK (Laskowski et al., 1993) shows that 96.7% of residues lie in the most favoured and additionally allowed regions, 2.8% are present in the generously allowed region, and 0.5% in the disallowed area. The four residues that fall in the disallowed region are 98 (monomer A), 464 (monomer A), 32 (monomer B) and 41 (monomer B).

The average temperature factor the for the room temperature nitrite reductase dimer is 61 Å$^2$. Temperature factors for the heme groups and protein atoms on a domain basis are given in Table 6.28.

**Table 6.28** Average temperature factors (Å$^2$) for the room temperature nitrite reductase

| Atoms | Monomer A | | Monomer B | |
|---|---|---|---|---|
| | N-terminal domain | C-terminal domain | N-terminal domain | C-terminal domain |
| main-chain | -- | 44 | 107 | 67 |
| side-chain | -- | 50 | 99 | 70 |
| *c* heme | - | | 98 | |
| $d_1$ heme | 42 | | 47 | |

Chapter 7

# *P. stutzeri* Nitrite Reductase: Structure and Function

## 7.1 INTRODUCTION

Though the current model of the *P. stutzeri* nitrite reductase is not complete it is still possible to make a number of conclusions about the structure and relate this to the function of this enzyme. The coordinates of the $d_1$ heme domain, and the $d_1$ heme itself, are well defined and allow the structural characterization of this domain. Although the *c* heme-binding domain is disordered, particularly in the frozen crystals, it has been possible to determine the overall topology of this domain and identify the *c* heme-binding ligands.

## 7.2 OVERVIEW OF THE STRUCTURE

The *P. stutzeri* nitrite reductase is organized into a functional dimer, with each monomer comprising two domains. The N-terminal *c* heme-binding domain is an all α-helical cytochrome *c*-like structure and the C-terminal, $d_1$ heme-binding, domain is an eight-bladed β-propeller structure. These two domains are folded separately and sequentially, with the N-terminal domain comprising residues 1-99 and the C-terminal domain comprising residues 133-533. The two are joined by a linker peptide, residues 100-132, which joins the two and then lies across the upper surface of the β-propeller domain. The dimer is formed by the interaction of the two β-propeller domains, with no contacts existing between the N-terminal domains. The crystallographic asymmetric unit comprises a dimer, which when measured along axes orthogonal to the two-fold NCS rotation axis, has approximate dimensions of 54 x 63 x 96 Å. The crystal lattice is dominated by interactions between the β-propeller domains, with the N-terminal domains contributing few crystal contacts.

Ribbon diagrams of two views of the *P. stutzeri* nitrite reductase dimer are presented in Figure 7.1 and the full amino acid sequence, with associated secondary structure assignments, is in Figure 7.2. During the course of the structural analysis some corrections were made to the original published sequence (Smith and Tiedje, 1992); the sequence given in Figure 7.2 contains these corrections, made as a result of the crystallographic analysis.

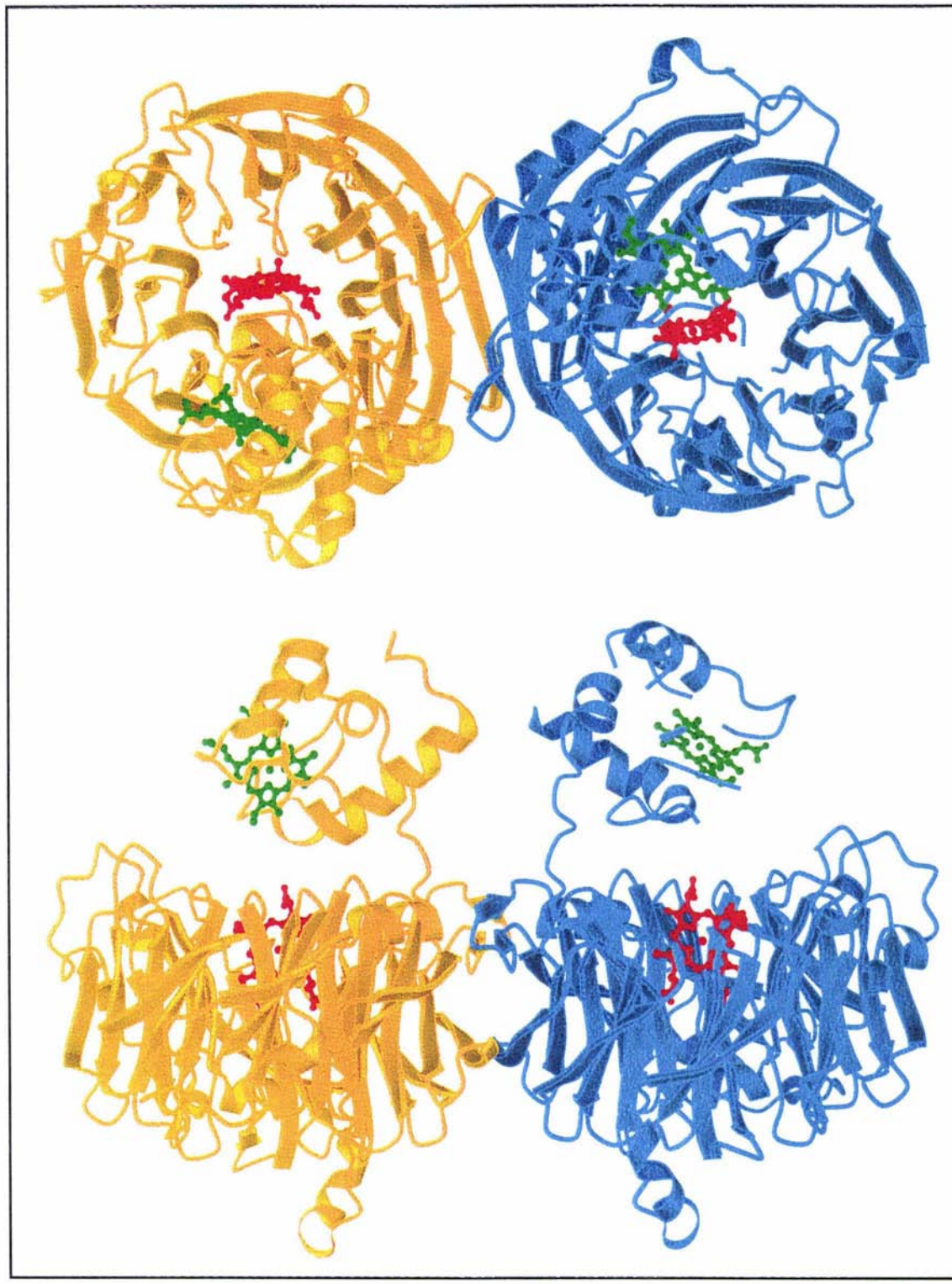


**Figure 7.1** Ribbon diagrams of two orthogonal views of the *P. stutzeri* NIR dimer.

The *c* (green) and $d_1$ (red) hemes are displayed in ball and stick representation. This figure was prepared with MOLSCRIPT (Kraulis, 1991).

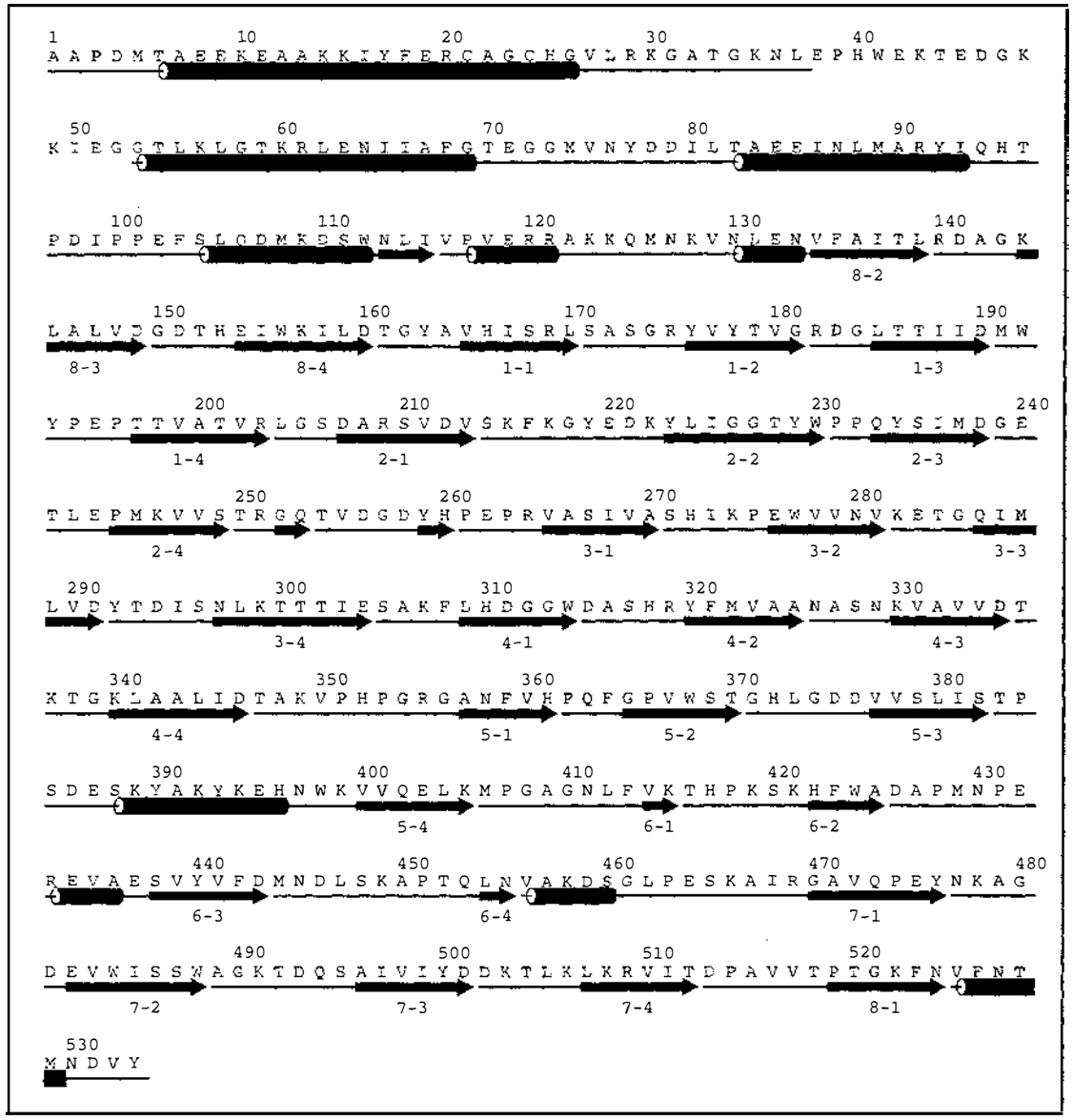


**Figure 7.2** Revised sequence of *P. stutzeri* JM300 NIR derived from the X-ray structure

The β-strands of the β-propeller domain are labelled such that the first digit refers to the propeller blade in which the strand is located, and the second digit corresponds to the position of the β-strand (1 for the innermost, 4 for the outermost) in the propeller blade. This figure was prepared with ALSCRIPT (Barton, 1993).

## 7.3 β-PROPELLER DOMAIN

### *7.3.1 Structural organisation*

The $d_1$ heme-binding domain (residues 133-533) forms an eight-bladed β-propeller structure. Each blade of the propeller comprises an anti-parallel β-sheet formed by

four strands. The β-sheets of each blade are twisted, and arranged radially around a central tunnel. The $d_1$ heme sits at the top of this channel in the centre of the β-propeller fold. A helix at the C-terminus of this domain folds up inside the other end of the central tunnel, closing it. The propeller structure is stabilized both by the β-sheets packing face to face, providing a framework for their predominately hydrophobic interactions, and by closing the β-propeller structure with the N and C-termini of the domain.

In the numbering scheme adopted for the β-propeller domain the β-strands are each referred to by two numbers. The first gives the number of the propeller blade in which the β-strand is located, with the second describing its position within the blade, 1 for inner-most and 4 for outer-most. By convention, the numbering of the propeller blades is such that closure is assigned to the last blade.

In the nitrite reductase β-propeller domain, the N-terminus provides the three outer-most strands of blade eight (8-2, 8-3, 8-4) before the chain folds to form blade one. The eighth blade is completed with the innermost β-strand (8-1) that comes from the C-terminus of the domain, just before the C-terminal α-helix (Figure 7.3).

The propeller β-strands are connected by loops and secondary structure elements, both between and within the blades. Two small β-strands involved in the dimer interface are inserted between blades two and three. A three turn $3_{10}$-helix (residues 386-396) is found between the third and fourth strand of blade five. A short α-helix (residues 432-438) is located between strands two and three of blade six, with a $3_{10}$-helix (residues 456-461) situated between blades six and seven.

Various attempts have been made to determine consensus sequences for the blades of β-propeller structures in order to identify these folds from sequence information alone e.g. the prediction that the α subunit of the integrins and the C-terminal region of phosphatidylinositol phospholipase D contain seven-bladed β-propeller folds (Springer, 1997). Examples of consensus sequences for β-propeller structures described to date include the "tryptophan docking" motif identified for the pyrroloquinoline quinone-binding proteins (Ghosh et al., 1996), the "Asp box" of the non-viral sialidases (Roggentin et al., 1989), and the "WD repeat" of the G-protein β subunit (Neer and Smith, 1996). The only repeating sequence unit found for the β-propeller domains from the nitrite reductase family is a conserved aspartate residue that is situated at the C-terminal end of the third strand of each propeller blade, with the

exception of blade 5 where the aspartate residue is located in the loop between strands 3 and 4 (Baker et al., 1997). This motif is conserved for the *P. stutzeri* nitrite reductase β-propeller domain, though the outlying aspartate of blade 5 is not present. These seven aspartate residues (residue numbers 149, 190, 238, 291, 335, 443, and 501) are involved in Asx turns (as described by Rees et al., 1983) between β-strands three and four of the propeller blade. In blade 5 an α-helix is inserted between these β-strands.

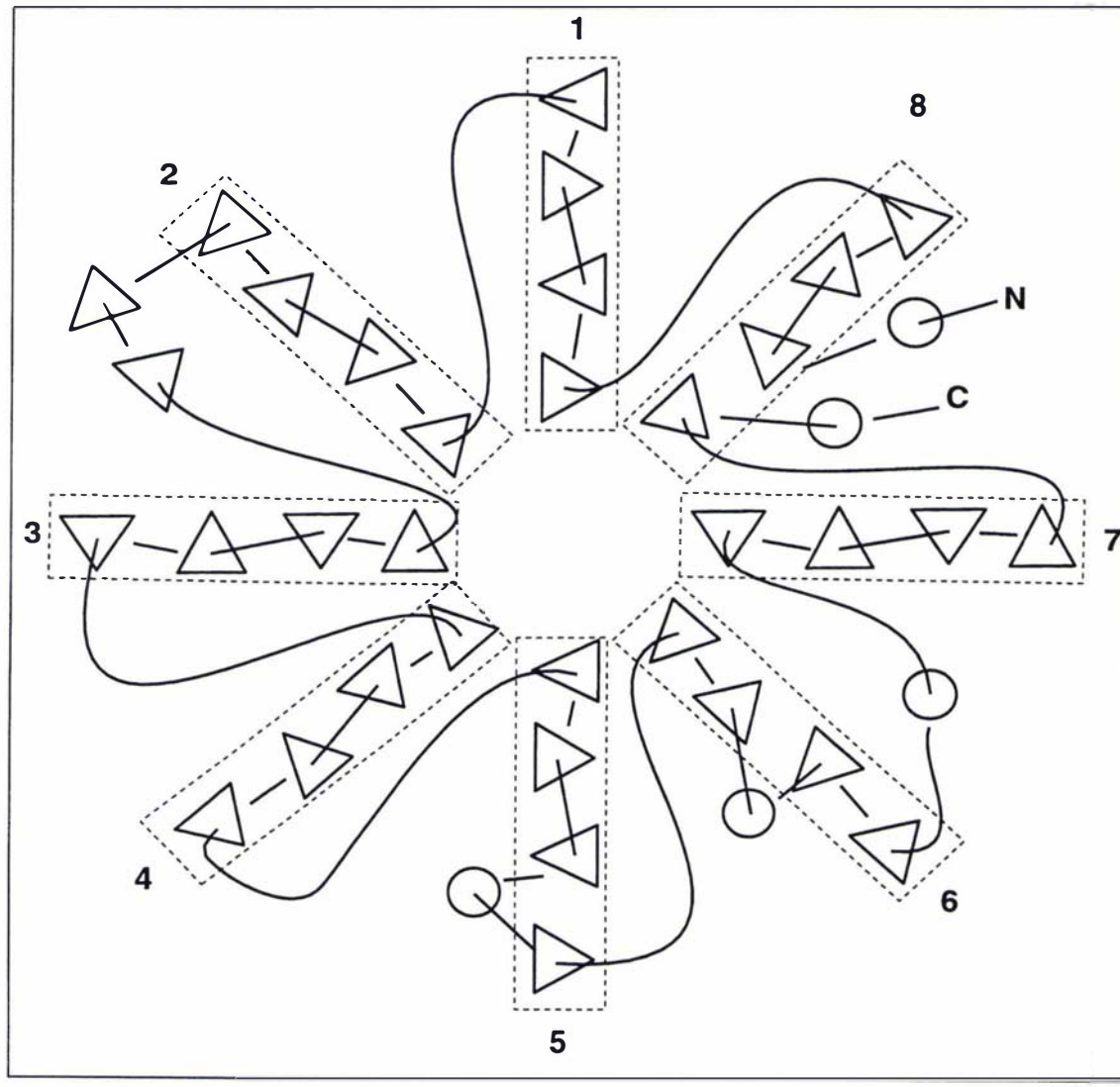



**Figure 7.3** Topology diagram of the β-propeller domain from *P. stutzeri* NIR

The β-propeller blades are boxed and numbered.

*7.3.1.1 Comparisons with the P. aeruginosa and T. pantotropha nitrite reductase β-propeller domains*

The conformation of the $d_1$ heme-binding domain for *P. stutzeri* nitrite reductase is very similar to that found for the homologous enzymes from *P. aeruginosa* and *T. pantotropha*. At residues 388-396, a five-residue insertion, together with amino acid substitutions either side, results in the insertion of a short $3_{10}$-helix between strands three and four of propeller blade five. A five-residue deletion between residues 448 and 449 results in a shortening of the loop between strands 6-3 and 6-4 relative to the *T. pantotropha* structure. This deletion is also observed for *P. aeruginosa* nitrite reductase. An insertion of residues 466-467 occurs in the *P. stutzeri* enzyme that increases the length of the loop immediately after the helix between blades six and seven. The only other insertion/deletion in the *P. stutzeri* $d_1$ heme-binding domain is a single amino acid deletion (between residues 257-258) in the β-hairpin that separates blades two and three and is involved in dimer contacts. The rms differences between the *P. stutzeri* β-propeller domain and linker region (residues 100-533), and the corresponding domains from *T. pantotropha* (Fulop et al., 1995) and *P. aeruginosa* (Nurizzo et al., 1997) nitrite reductases are 0.85 Å (for 411 Cα atoms) and 0.88 Å (for 423 Cα atoms) respectively. All Cα atoms conserved between these domains were used in the calculation. These domains are shown superimposed in Figure 7.4.

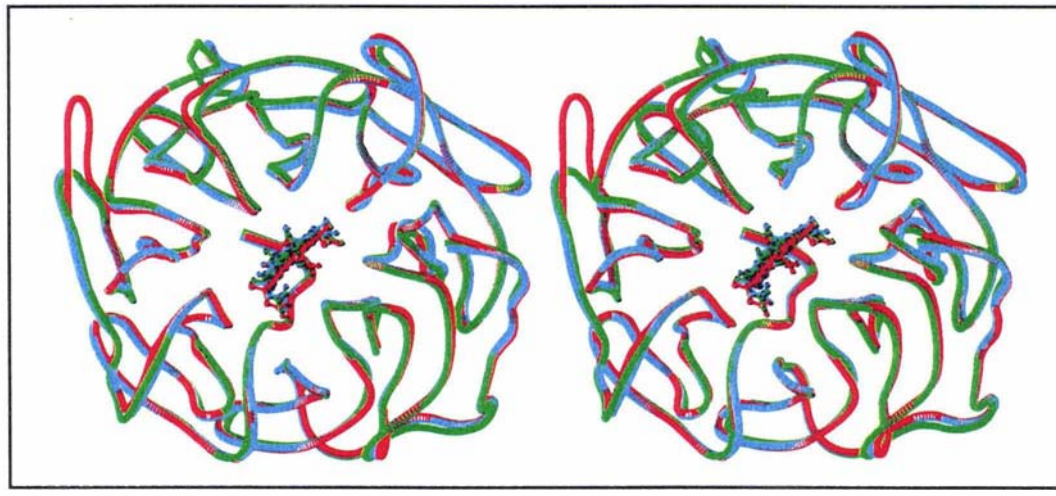


**Figure 7.4** Stereo Cα superposition of the *P. stutzeri* (green), *P. aeruginosa* (blue) and *T. pantotropha* (red) NIR β-propeller domains

The superposition was calculated on all Cα atoms conserved between the NIRs from the three species. The $d_1$ heme groups are shown in ball and stick representation. This figure was prepared with MOLSCRIPT (Kraulis, 1991).

*7.3.1.2 Comparisons with other β-propeller structures*

Crystal structures have been determined for various proteins containing β-propeller folds that are comprised of between four and eight blades. Four-bladed domains have been observed for hemopexin (Faber et al., 1995) and collagenase (Li et al., 1995), whereas influenza virus neuraminidase (Varghese et al., 1983) and bacterial sialidase (Crennell et al., 1993) exhibit six-bladed structures. The seven-bladed β-propeller structures are the most numerous examples of this folding pattern to date, with members of this group including methylamine dehydrogenase (Vellieux et al., 1989), galactose oxidase (Ito et al., 1991), the G-protein β subunit (Wall et al., 1995) and the RCC1 regulator of chromosome condensation (Renault et al., 1998). The only eight-bladed β-propeller fold other than that of the nitrite reductase family to be structurally characterized is from the enzyme methanol dehydrogenase (Xia et al., 1992). Methanol dehydrogenase also binds its cofactor, pyrrolo-quinoline quinone (PQQ), in the centre and top of the β-propeller domain. This enzyme utilizes a different manner for 'tying up' the propeller structure, with the N-terminus providing only the outermost strand of blade eight, with the three inner strands provided by the C-terminus. A superposition of 126 Cα atoms from propeller blades 1 to 7 gives an rms difference of 3.6 Å between the *P. stutzeri* NIR and *Methylophilus* W3a1 methanol dehydrogenase eight-bladed β-propeller structures. These structures are shown superimposed in Figure 7.5.

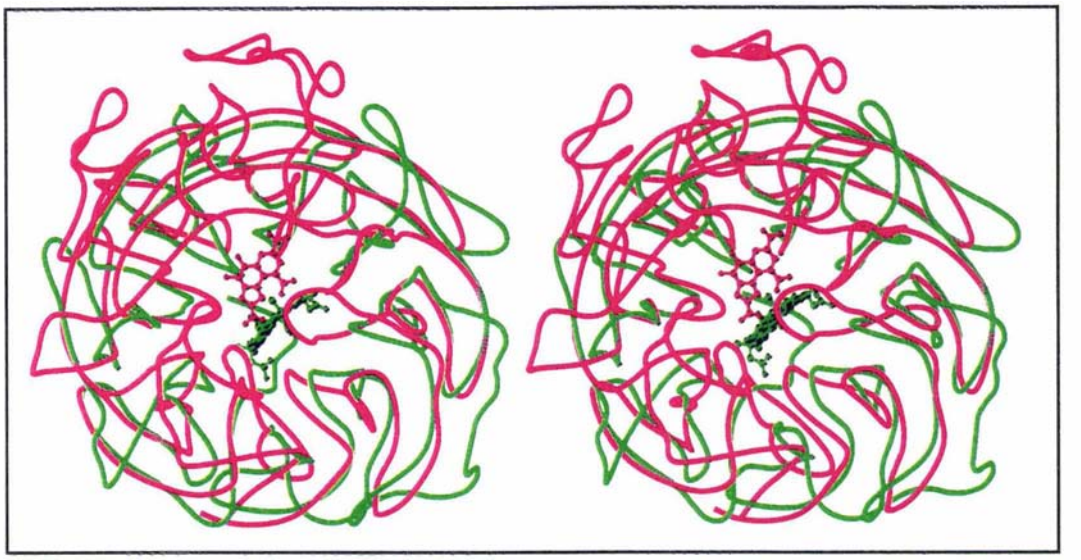


**Figure 7.5** Stereo Cα superposition of the β-propeller structures of *P. stutzeri* NIR (green) and *Methylophilus* W3a1 methanol dehydrogenase (purple)

The N-terminal domain of NIR has been omitted for clarity. The $d_1$ heme and PQQ cofactors are shown in ball and stick representation. This figure was prepared with MOLSCRIPT (Kraulis, 1991).

### *7.3.2 $d_1$ heme environment and structure*

The $d_1$ heme sits in a pocket formed in the top and centre of the β-propeller domain. The heme group is bound by an axial histidine ligand (residue 166) which is provided by the innermost strand of blade one (1-1), with an Fe-Nε2 bond length of 2.1 Å. The other axial ligand is a solvent molecule, presumed to be water, which is clearly distinguished in the $m|F_o| - D|F_c|$ map as a four sigma peak on all four $d_1$ hemes characterized (two frozen monomers and two room temperature monomers) (Figure 7.7). This water molecule is bound with an Fe-O distance of 2.4 Å. The heme pocket is predominately hydrophobic, owing to the presence of residues Ile 167, Leu 373, Leu 412, Phe 413, Met 429, Val 472, and Phe 523. Five conserved arginine residues (140, 169, 182, 209, 355) provide hydrogen bonding partners for the propionate and carbonyl groups of the $d_1$ heme (see Table 7.1 and Figure 7.6). It was the apparent absence of one of these arginines, and the presence of density characteristic of an arginine interacting with the heme propionate, that helped identify one of the regions of error in the published sequence for this enzyme.

**Table 7.1** Potential hydrogen bonds to the $d_1$ heme

| Heme Atom | Contact Atom | Distance (Å) |
|---|---|---|
| O1A | 268 Ile N | 3.6 |
| O2A | 211 Val N | 3.6 |
| O1B | 182 Arg NH2 | 2.4 |
| O2B | 182 Arg NH2 | 2.4 |
| O2B | 209 Arg NE | 3.2 |
| O1C | 140 Arg NE | 3.0 |
| O2C | 140 Arg NE | 3.6 |
| O1D | 355 Arg NE | 3.5 |
| O1D | 355 Arg NH2 | 3.6 |
| O2D | 355 Arg NE | 2.7 |
| O2D | 355 Arg NH2 | 3.1 |
| O2D | 169 Arg NH2 | 3.4 |
| OMB | 229 Tyr OH | 2.9 |
| OMB | 209 Arg NE | 3.0 |

The numbering convention for the $d_1$ heme is given in Baker et al., 1997.

Two conserved histidine residues (310 and 352) are present on the distal, substrate-binding, side of the $d_1$ heme in positions in which they could hydrogen bond to the substrate (Figure 7.6).

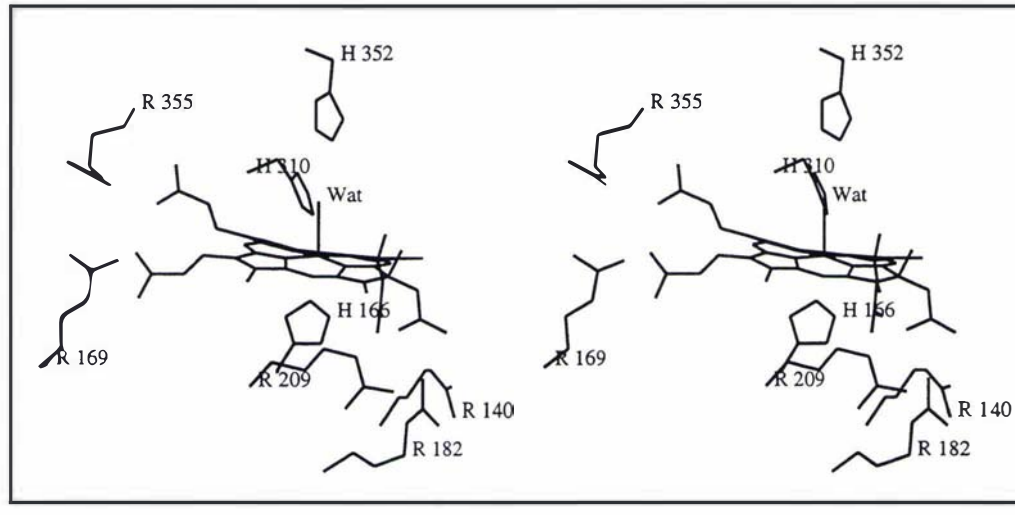



**Figure 7.6** Stereo view of the $d_1$ heme group of *P. stutzeri* NIR

A water molecule is shown bound to the $d_1$ heme iron. This figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

The most striking feature of the $m|F_o| - D|F_c|$ and $2m|F_o| - D|F_c|$ electron density maps of the $d_1$ heme pocket is the presence of a very large electron density peak extending from the edges of the B and C pyrrole rings. This peak is found for both monomers in both the room temperature and frozen crystals, though the largest peaks are present for the room temperature data. The largest peak of this extra density is found for monomer A in the room temperature lattice at a height of 6 sigma above the mean, located 15 Å away from the edge of the $d_1$ heme. The electron density is continuous from the edge of the heme to this large peak, though at a lower contour level. This difference density, along with a peak for a water ligand omitted from the map calculation, is shown in Figure 7.7. Curiously, a portion of this density occupies a similar position to that of Tyr 10 in the *P. aeruginosa* nitrite reductase (see Section 7.4.2), with the rest of this density serving to close off the heme pocket in a similar manner to the N-terminal region of the *P. aeruginosa* and *T. pantotropha* enzymes (Figure 7.8). Remarkably, it is possible to fit the large alkyl side-chain of an *a* heme into this difference density by rotating about the side-chain torsion angles. Unfortunately, owing to the limited resolution of the data it is not possible to be certain of whether the density originates from the $d_1$ heme, or what the nature of this substituent might be.

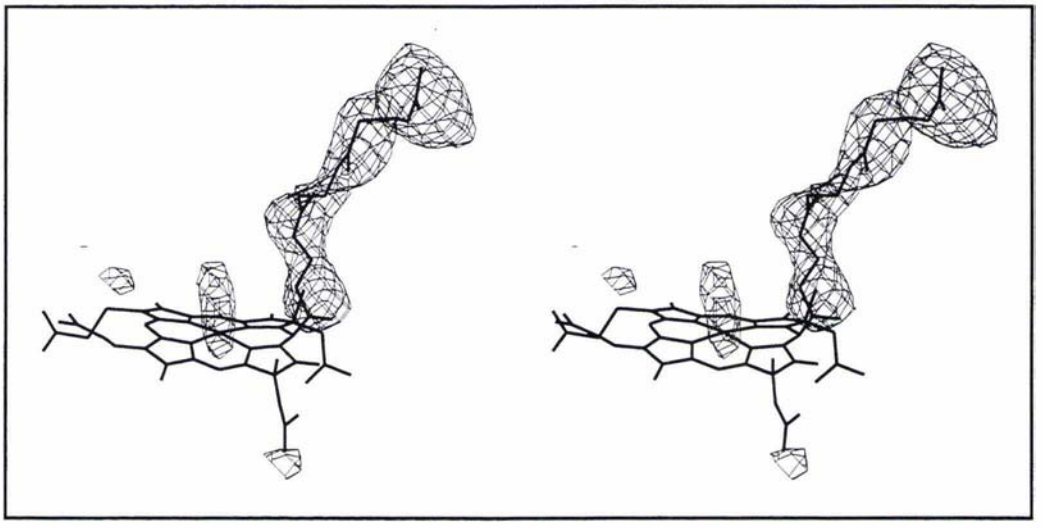


**Figure 7.7** Stereo view of the $m|F_o| - D|F_c|$ map of the $d_1$ heme showing the unaccounted for electron density and modelling of the heme *a* side-chain

The map is contoured at 3σ, and the *a* heme side-chain is overlaid. This figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

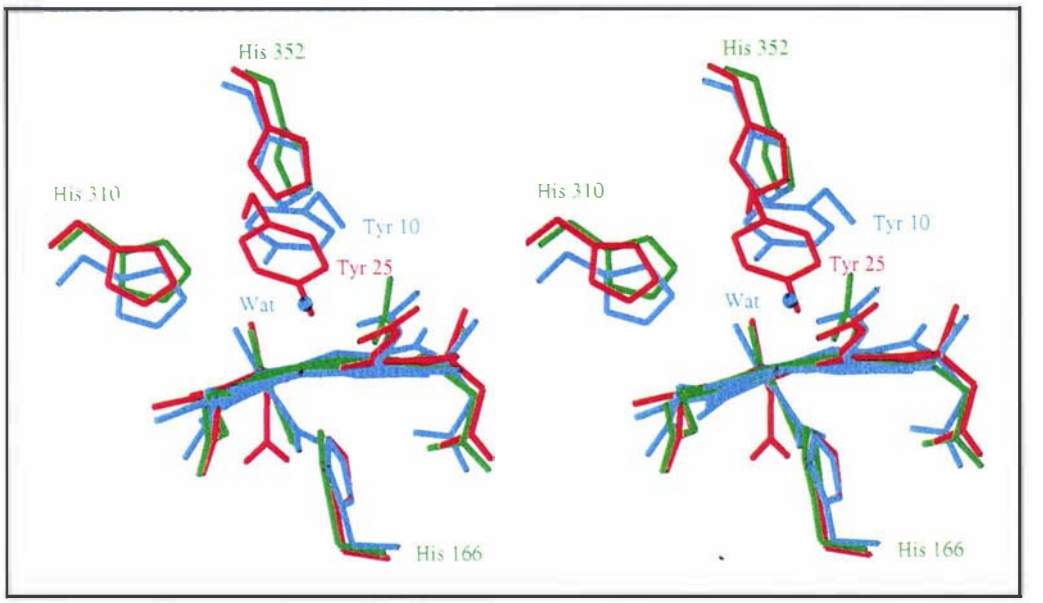



**Figure 7.8** Stereo view of the *P. stutzeri* (green), *P. aeruginosa* (blue) and *T. pantotropha* (red) NIR $d_1$ heme groups.

The superposition was calculated on all atoms of the $d_1$ heme. This figure was prepared with MOLSCRIPT (Kraulis, 1991).

If an *a* heme side-chain is modelled into the density, the model refines to a free R-factor of 0.287 with an R-factor of 0.235 (compared to values of 0.289 and 0.236 without this side-chain), with a grouped temperature factor for the side-chain of 70 Å$^2$. This suggests that the density could be due to an alkyl group of some kind. However, this side-chain does not extend far enough away from the $d_1$ heme and does not fully account for the large difference peak located 15 Å away from the heme edge when included in the Fourier synthesis. It has been shown that the $d_1$ heme can be removed from the *P. stutzeri* and *P. aeruginosa* nitrite reductases (Hill and Wharton, 1978; Walsh et al., 1980; Weeg-Aerssens et al., 1991) and replaced with either native or synthetically prepared $d_1$ heme. The activity of the *P. stutzeri* enzyme is reduced to 82% and 77% of the native protein on re-incorporation of the native or synthetic $d_1$ heme respectively, implying that the proposed structure for heme $d_1$ is correct (Weeg-Aerssens et al., 1991). The high resolution structures of the *T. pantotropha* and *P. aeruginosa* NIRs have confirmed the identity of the $d_1$ heme in the enzymes from these species. The spectral properties of the reconstituted enzyme are indistinguishable from the native enzyme. Another possibility for the origin of this peak is that it may correspond to an Fe atom of the disordered *c*

heme domain in a partially occupied orientation required to transfer an electron to the $d_1$ heme. However, this peak is closer (15 Å) than the *c* heme to $d_1$ heme distances found for the monomers of *T. pantotropha* (20.6 Å) and *P. aeruginosa* (19.6 Å), and there exists no surrounding density that could be ascribed to any polypeptide. When an iron atom was positioned in the centre of this peak and refined, its temperature factor reached a value of 94 $Å^2$, with the model having free and conventional R-factors of 0.287 and 0.235 respectively. A further possible explanation is that the extra density observed could be part of the twelve-residue insertion (residues 45-56, see Figure 5.4) in the *P. stutzeri* NIR. The distance between the region of the N-terminal domain in which this insertion occurs and the $d_1$ heme is 18-20 Å, so it is possible for the insertion to connect to this region. However, no electron density is visible extending away from the N-terminal domain.

Density is apparent in a similar position for the $d_1$ heme of the second room temperature monomer; this density is not continuous at the 2σ level, but it does extend further away. A peak is also visible for the frozen lattice NIR, but is not present at such a high level. Unfortunately, owing to the poor resolution and quality of the experimental data, it is not possible to identify the origin of this extra density peak or to state whether it is a substituent of the heme or is a part of the polypeptide that is in close proximity. What is clear, is that this scattering matter functions to close off the heme pocket in a similar manner to the N-terminal extension found for the homologous enzymes from other species. Though extensive functional measurements have not been conducted on the nitrite reductase from *P. stutzeri*, exposure of the $d_1$ heme to solvent would be expected to have dramatic effects on the redox potential of the heme group and the activity of the enzyme.

Attempts to obtain a molecular weight by mass spectrometry for the $d_1$ heme group removed from the *P. stutzeri* nitrite reductase by extraction into acidic-acetone (as described by Walsh et al., 1980) have been unsuccessful to date.

## 7.4 N-TERMINAL DOMAIN

### *7.4.1 Structural organisation*

Though parts of the N-terminal domain have not been modelled (see Section 6.7.14), and the domain appears disordered in the lattice, some preliminary conclusions can be drawn from the current models. The N-terminal domain (residues 1-99) is an all α-helical domain similar to the class I *c*-type cytochromes, e.g.

cytochrome $c_{551}$ from *P. stutzeri* (Cai et al., 1992) and *P. aeruginosa* (Matsuura et al., 1982; Timkovich and Cai, 1993), or tuna cytochrome *c* (Takano and Dickerson, 1981) (Figure 7.9). The domain is comprised of 4 α-helices formed by residues 11-20, 22-25, 59-67 and 89-94. The *c* heme is covalently bound by thioether linkages to Cys 21 and Cys 24 with His 25 providing an axial ligand to the heme iron (Figure 7.10). These residues comprise the *c* heme binding motif C-X-X-C-H common to the *c*-type cytochromes. The second axial heme ligand is provided by Met 74. This histidine-methionine coordination of the *c* heme is the classical mode of coordination for the class I *c*-type cytochromes. The observation of this binding mode confirms the spectroscopic studies conducted on the oxidised *P. stutzeri* ZoBell nitrite reductase that indicated the presence of histidine-methionine coordination of the *c* heme (Cheesman et al., 1997).

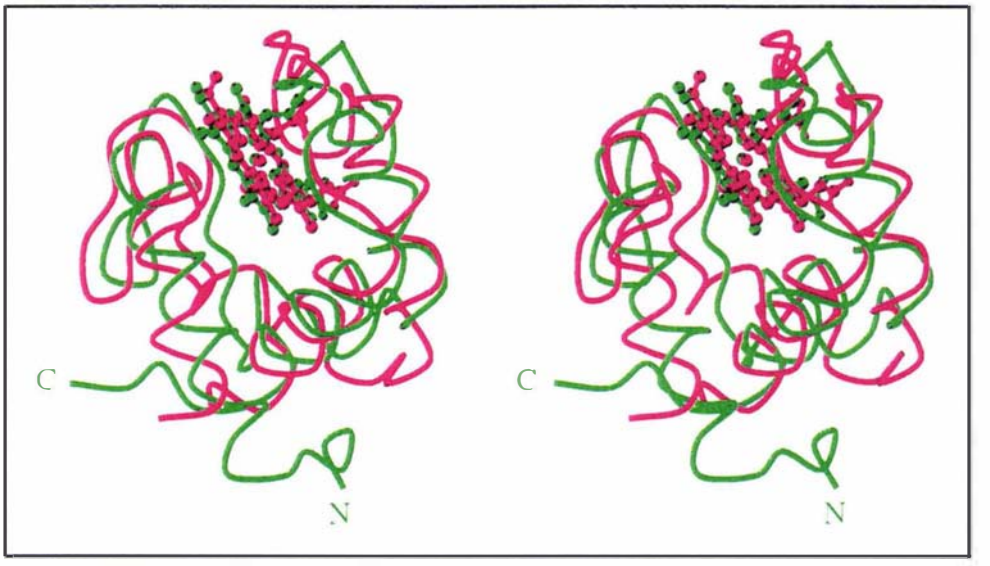



**Figure 7.9** Stereo Cα trace of a superposition of the NIR N-terminal domain (green) and cytochrome $c_{551}$ (purple) from *P. stutzeri*

The *c* heme groups are shown in ball and stick representation. The figure was prepared with MOLSCRIPT (Kraulis, 1991).

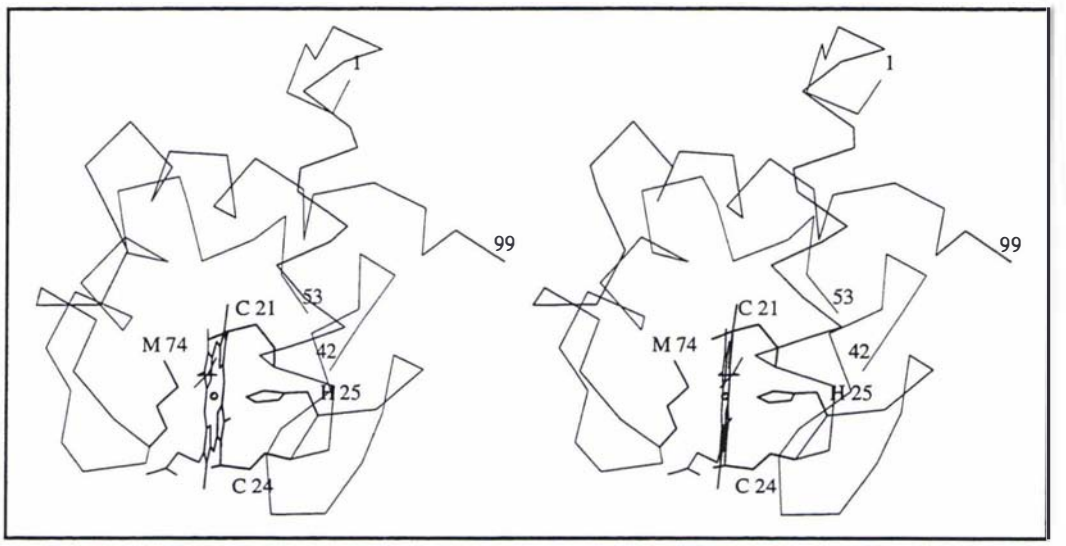



**Figure 7.10** Stereo view of the *P. stutzeri* NIR N-terminal domain and *c* heme ligands

This figure was prepared with MOLSCRIPT (Kraulis, 1991).

### *7.4.2 Comparisons with the N-terminal domains from T. pantotropha and P. aeruginosa nitrite reductases*

The N-terminal domain of *P. stutzeri* nitrite reductase is very similar in overall structure to the comparable domains found for the oxidised *P. aeruginosa* (Nurizzo et al., 1997) and reduced *T. pantotropha* enzymes (Williams et al., 1997). The most striking difference is at the N-terminus, where the *P. stutzeri* enzyme has a large deletion relative to both the *P. aeruginosa* and *T. pantotropha* enzymes (see the sequence alignment in Figure 5.4). This has important implications for the $d_1$ heme iron coordination, and for connections between the two domains. The N-terminal deletion of 26 residues, compared with the *P. aeruginosa* NIR sequence, results in the removal of a tyrosine residue (residue 10) which hydrogen bonds to the water ligand of the $d_1$ heme iron atom in *P. aeruginosa* NIR. The *P. stutzeri* NIR N-terminal domain is 52 residues shorter than that of *T. pantotropha* NIR. This deletion again results in the removal of a tyrosine (residue 25), which in this case is coordinated directly to the oxidised $d_1$ heme iron (Figure 7.11).

Unlike the oxidised NIRs from *P. stutzeri* and *P. aeruginosa* which coordinate the *c* heme iron atom with histidine and methionine residues, the oxidised *T. pantotropha* enzyme exhibits bis-histidine coordination (Fulop et al., 1995). When the *T. pantotropha* nitrite reductase is reduced, the *c* heme-binding domain undergoes a struc-

tural rearrangement that results in the N-terminal histidine ligand of the *c* heme being replaced by a methionine, to give the same ligation mode found for the oxidised *P. stutzeri* and *P. aeruginosa* NIRs. This structural rearrangement also results in Tyr 25 no longer coordinating the $d_1$ heme iron atom, increasing the solvent accessibility of the heme and leaving a vacant coordination site to enable substrate binding (Williams et al., 1997).

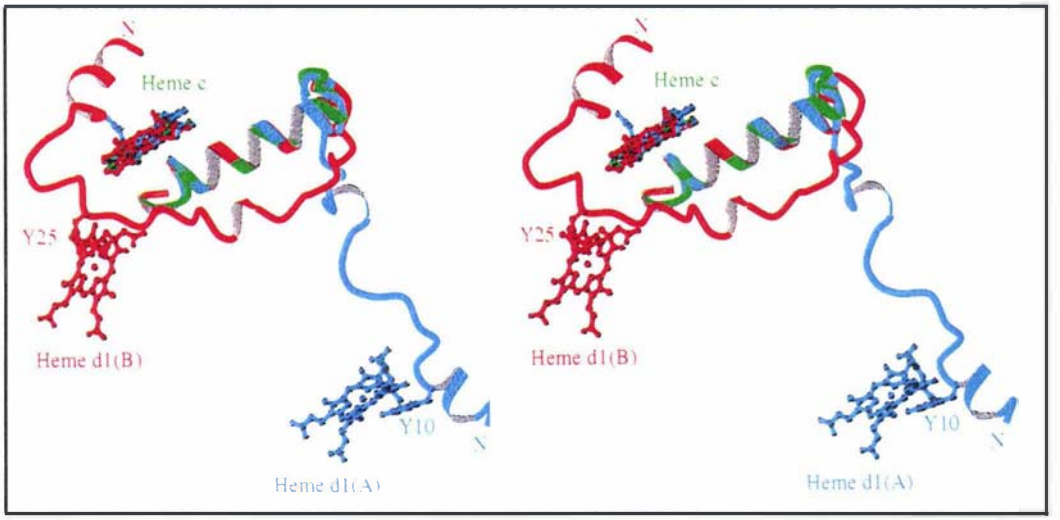



**Figure 7.11** Structural diversity of the N-terminal extension for the three oxidized NIRs

The N-terminal region and *c* heme of monomer B from *P. stutzeri* NIR (green), *P. aeruginosa* NIR (blue) and *T. pantotropha* NIR (red) are shown. Interactions of the $d_1$ heme with the tyrosine side-chains of the latter two enzymes are highlighted. The figure was prepared with MOLSCRIPT (Kraulis, 1991).

The large twelve-residue insertion (residues 45-56) in the N-terminal domain of the *P. stutzeri* enzyme, relative to the domains of the *T. pantotropha* and *P. aeruginosa* enzymes is not visible in the electron density maps at this stage, so it is not clear what effect this has on the overall fold of the domain. This insertion is in a similar position in the sequence to the 15-residue insertion found for the mitochondrial *c*-type cytochromes (e.g. tuna cytochrome *c*, residues 41-55) relative to the bacterial proteins. However, no sequence similarity exists between these two polypeptide insertions. Secondary structure prediction using the profile network method implemented in PHD (Rost and Sander, 1994a; Rost and Sander, 1994b) predicted that this insertion forms a loop with a relatively high solvent accessibility. This result was consistent when the algorithm was run either on the isolated fragment, or on the whole N-terminal domain; in the latter case the algorithm did predict the known helices of this fold.

Apart from the N-terminal region, the only other deletion that occurs for the *P. stutzeri c* heme-binding domain is a two residue deletion that occurs between residues 77 and 78, relative to the *P. aeruginosa* and *T. pantotropha* nitrite reductases. This two residue deletion results in little change in the overall fold of the domain.

## 7.5 RELATIONSHIP OF THE *c* HEME-BINDING AND $d_1$ HEME-BINDING DOMAINS

### *7.5.1 Structural organisation of the monomer*

As noted earlier, the nitrite reductase monomer from *P. stutzeri* is comprised of two domains; an all α-helical N-terminal domain (residues 1-99) that binds the *c* heme and has a fold resembling the class I *c*-type cytochromes, and the C-terminal, $d_1$ heme-binding, eight-bladed β-propeller domain (residues 133-533). The two domains are connected via a short linker region (residues 100-132). A Cα-trace of the *P. stutzeri* nitrite reductase monomer showing the relationship of the two domains is given in Figure 7.12.

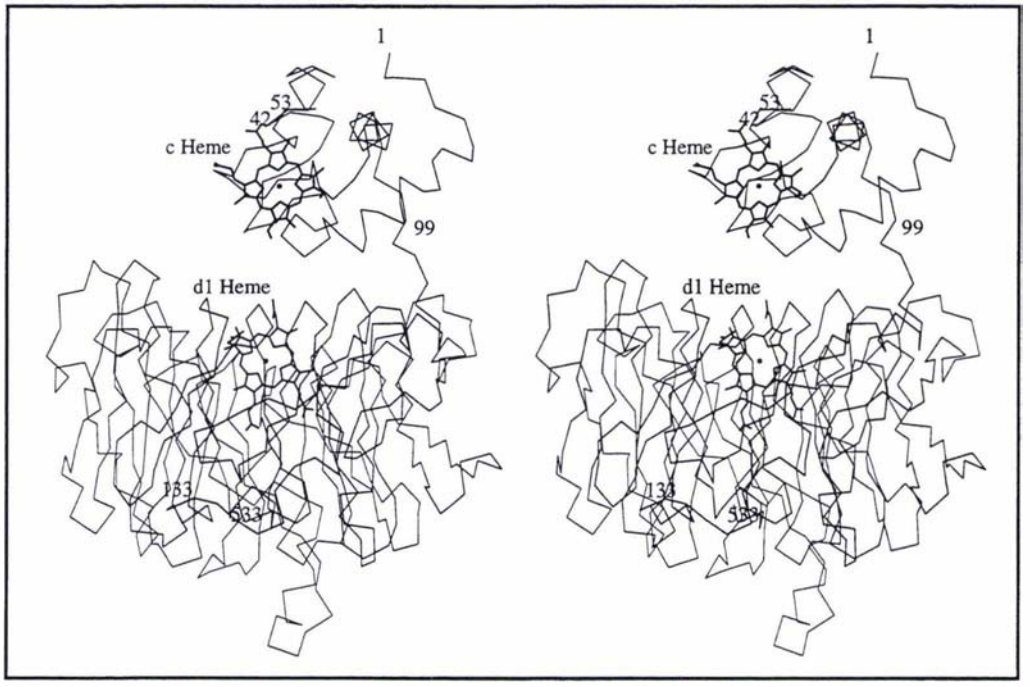



**Figure 7.12** Stereo view of the *P. stutzeri* NIR monomer

The figure was prepared with MOLSCRIPT (Kraulis, 1991).

Examination of the Fe...Fe distances within each nitrite reductase monomer (Table 7.2) shows that the positions of the redox centres are consistent with the hypothesis that each monomer functions independently in electron transport.

**Table 7.2** Relationship of the *c* and $d_1$ heme groups within the NIR monomers

| NIR subunit | $d_1$ heme - *c* heme angle (°)[a] | Fe...Fe distance (Å)[b] | $d_1$ - *c* heme edge distance (Å) |
|---|---|---|---|
| monomer A | 55 | 24.1 | 19.4 |
| monomer B | 27 | 26.7 | 22.2 |

[a] Angle between the two heme planes, calculated using all heme atoms.

[b] Distance between $d_1$ heme iron and *c* heme iron.

The distances between the heme redox centres of different monomers are much greater and seem incompatible with electron transfer between the subunits of the dimer. The distances between redox centres contained within the two monomers are listed in Table 7.3.

**Table 7.3** Intermolecular distances between the iron atoms of the two monomers

| Redox centres | Fe...Fe distance (Å) |
|---|---|
| monomer A $d_1$ Fe - monomer B $d_1$ Fe | 49.7 |
| monomer A *c* Fe - monomer B $d_1$ Fe | 54.4 |
| monomer A $d_1$ Fe - monomer B *c* Fe | 54.1 |
| monomer A *c* Fe - monomer B *c* Fe | 52.3 |

### *7.5.2 Non-crystallographic symmetry relationship of the P. stutzeri monomers*

A least squares superposition of the two monomers that comprise the NIR dimer reveals that the two N-terminal domains, and the two C-terminal domains, are related by different non-crystallographic symmetry operators. The C-terminal β-propeller domains are related by a true two-fold rotation. The two N-terminal domains, however, are related by a 170° rotation about a different axis to that of the β-propeller domains (Table 7.4).

**Table 7.4** NCS transformations of the *P. stutzeri* NIR dimer

| NIR domain | Rotation (spherical polar angles) (°) | | | Translation vector (Å) | | |
|---|---|---|---|---|---|---|
| | ω | ϕ | χ | x | y | z |
| N-terminal | 87 | -44 | 170 | 36.0 | 39.1 | -38.6 |
| C-terminal | 92 | 119 | 179 | 45.9 | 24.3 | -31.3 |

The spherical polar angles are defined such that ω is the angle from the z axis and ϕ is the angle from the x axis to the y axis, with χ being the rotation about the axis defined by ω and ϕ. The superpositions were calculated on all Cα atoms.

The overall effect of this different rotation axis for the N-terminal domain, is that when the monomers are superimposed on the β-propeller domains, an additional 36° rotation (about the axis ω=20 ϕ=135) and translation (by the vector 15.2, 0.3, 3.2 Å) are required to overlay the two N-terminal domains. If the N-terminal *c* heme-binding domains obeyed the same NCS operator as the β-propeller domains then unfavourable steric clashes would occur in the crystal lattice.

### *7.5.3 Comparisons with the structures of P. aeruginosa and T. pantotropha nitrite reductases.*

The N-terminal domains of *P. stutzeri* NIR have undergone a large rigid-body rotation, relative to the β-propeller domain, when compared to the structures for *P. aeruginosa* and *T. pantotropha* nitrite reductases. In order to calculate the transformations for the N-terminal domains from *P. stutzeri* NIR and the homologous NIRs, the monomers were initially superimposed on the β-propeller domains and then the transformation required to overlay the N-terminal domains was calculated. The transformations for the superposition of the *P. stutzeri* N-terminal domains on to the corresponding domains from the homologous NIRs are given in Table 7.5. The ~98° rotation of the *P. stutzeri* N-terminal domain of monomer B compared to the corresponding domains from *P. aeruginosa* and *T. pantotropha* NIRs is shown in Figure 7.13.

**Table 7.5** Transformations for superposition of the NIR N-terminal domains

| *P. stutzeri* NIR N-terminal domain | Corresponding NIR N-terminal domain | Rotation (°) (spherical polar angles) | | | Translation vector (Å) | | |
|---|---|---|---|---|---|---|---|
| | | ω | ϕ | χ | x | y | z |
| A | *P. aeruginosa* | 108 | 139 | 76 | 13.7 | 24.7 | 23.2 |
| | *T. pantotropha* ox. | 110 | 142 | 65 | 10.2 | 23.1 | 23.2 |
| | *T. pantotropha* red. | 105 | 137 | 54 | 9.5 | 16.8 | 23.2 |
| B | *P. aeruginosa* | 62 | 118 | 99 | 46.3 | 29.3 | -8.4 |
| | *T. pantotropha* ox. | 63 | 114 | 88 | 43.1 | 21.0 | -6.0 |
| | *T. pantotropha* red. | 68 | 120 | 96 | 44.8 | 29.0 | -8.1 |

The spherical polar angles are defined such that ω is the angle from the z axis and ϕ is the angle from the x axis to the y axis, with χ being the rotation about the axis defined by ω and ϕ. The superpositions were calculated on all conserved Cα atoms. ox. and red. are abbreviations for oxidized and reduced.

In the oxidised forms of the *P. aeruginosa* and *T. pantotropha* enzymes there are a substantial number of interactions between the *c* heme and $d_1$ heme-binding domains. In contrast, for the *P. stutzeri* NIR monomer there are no other interactions between the two domains apart from those involving the polypeptide linker region that connects them. With no such inter-domain interactions, and with the crystal contacts dominated by contacts between the β-propeller domains, it is perhaps not surprising that the N-terminal domains are mobile and can adopt a different conformation to that found for the *T. pantotropha* and *P. aeruginosa* monomers.

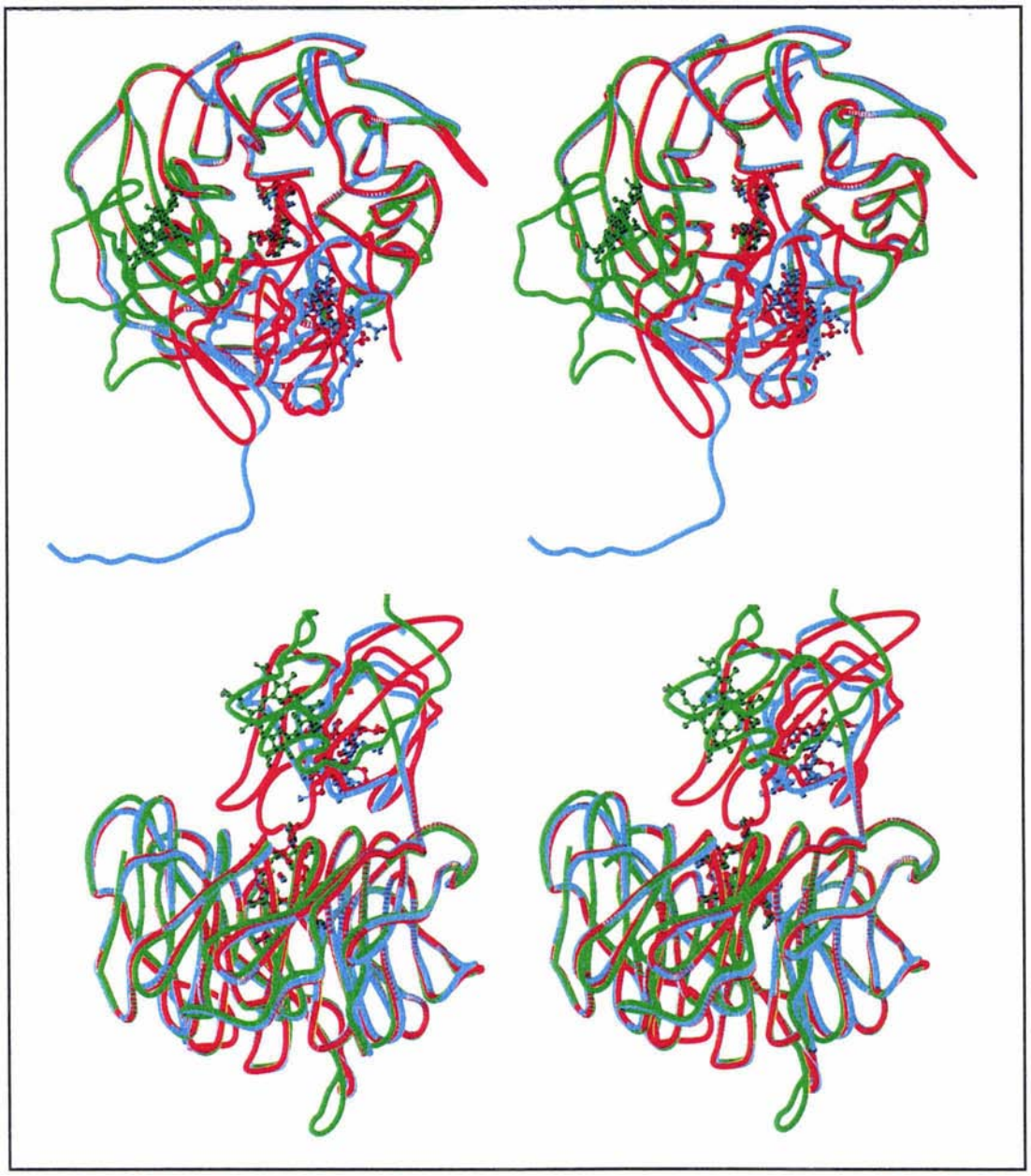


**Figure 7.13** Two orthogonal stereo Cα traces of the *P. stutzeri* (green), *P. aeruginosa* (blue) and *T. pantotropha* (red) NIR monomers.

The overlay was calculated on all conserved Cα atoms of the β-propeller domains. The N-terminal extension of the *P. aeruginosa* NIR has been truncated for clarity in the lower diagram. The figure was prepared with MOLSCRIPT (Kraulis, 1991).

In *P. aeruginosa* nitrite reductase the N-terminal arm (residues 1-29) interacts with the β-propeller domain of the other monomer in the dimer, in a process termed 'N-terminal arm exchange', or 'domain swapping' (Nurizzo et al., 1997). The N-terminal extension is responsible for forming a large proportion of the interactions between

the *c* heme and $d_1$ heme-binding domains. Apart from the interactions formed by the N-terminal extension there exist only five other hydrogen bonds and a single van der Waals contact between the *c* and $d_1$ domains, which presumably would not be sufficient to tether the domains together (Nurizzo et al., 1997). Silvestrini and co-workers have shown that the *c* and $d_1$ domains of *P. aeruginosa* nitrite reductase can be separated by cleavage with subtilisin. If the separated domains are placed in solution together, the enzyme is unable to oxidize its physiological substrate cytochrome $c_{551}$, implying that the two domains do not have a significant affinity for one another once the polypeptide linker is broken (Silvestrini et al., 1996).

For the oxidised *T. pantotropha* nitrite reductase most of the inter-domain interactions are provided by the N-terminal extension, though in this case it associates with the β-propeller domain of the same monomer. Apart from the interactions made by the N-terminal extension, seven other hydrogen bonds are formed between the two domains (Baker et al., 1997).

When the *T. pantotropha* nitrite reductase crystals are reduced, the *c* heme-binding domain undergoes a structural rearrangement (as described previously Section 7.4.2). This conformational change does not affect the integrity of the crystal lattice as the crystal contacts are dominated by interactions between the β-propeller domains. A consequence of this rearrangement is that the N-terminal extension of the protein no longer forms any interactions with the β-propeller domain. In one molecule of the dimer (monomer A) all the hydrogen bonding interactions are broken between the *c* and $d_1$ domains, whereas a third of these interactions are disrupted in the other monomer. For monomer A the first 35 residues are disordered, and residues 63-121 are only weakly visible in the electron density maps (and hence modelled at half occupancy). The first 25 residues are disordered in the second monomer. The relative positions of the N and C-terminal domains within the NIR dimer are also altered, with the two N-terminal domains and the two C-terminal domains no longer related by the same non-crystallographic symmetry axis. It is proposed that owing to the lack of crystal contacts involving the *c* heme domains, and the dramatic decrease in the number of interactions between the N and C-terminal domains of the monomer, the *c* heme-binding domains are free to move within the lattice.

A similar situation appears to exist for the *P. stutzeri* nitrite reductase where there is no N-terminal extension that can provide the large number of inter-domain contacts that occur for the homologous *P. aeruginosa* and *T. pantotropha* oxidised enzymes.

Calculation of the water-accessible surface buried at the N-terminal domain/C-terminal domain interface of the NIR monomers illustrates the dramatic effect of removing the N-terminal extension from this region of the structure, either by deletion as is the case for *P. stutzeri* NIR, or by structural rearrangement as is found for the reduced *T. pantotropha* enzyme (Table 7.6).

**Table 7.6** The water accessible surface buried between the NIR N-terminal and C-terminal domains

| NIR monomer | Buried surface (Å$^2$) |
|---|---|
| *P. stutzeri* (A) | 532 |
| *P. stutzeri* (B) | 294 |
| *T. pantotropha* red. (A) | 400 |
| *T. pantotropha* red. (B) | 1448 |
| *T. pantotropha* ox. | 2782 |
| *P. aeruginosa** | 2024 |

*The N-terminal domain of *P. aeruginosa* NIR was defined as the core (residues 29-115) of one N-terminal domain and the domain swapped N-terminal extension (residues 4-28) from the other monomer.

The buried surface is calculated for both monomers of the *P. stutzeri* and reduced *T. pantotropha* NIRs as the two N-terminal, and two C-terminal domains of the dimer obey different NCS operators, hence the relationship between the N and C-terminal domains is different for each monomer.

The solvent accessible surface was calculated using the programs AREAIMOL and RESAREA from the CCP4 suite (Collaborative Computational Project, 1994), with a solvent probe of radius 1.4 Å; ox. and red. are abbreviations for oxidized and reduced.

The substantially fewer inter-domain interactions, combined with the fact that crystal contacts are predominantly formed between the β-propeller domains, allows the *P. stutzeri* N-terminal domains freedom of movement within the lattice and contributes to their partial disorder.

## 7.6 DIMER INTERFACE

The *P. stutzeri* nitrite reductase dimer is formed by the association of the two β-propeller domains, with no contacts made between the two N-terminal domains. The dimer interface is formed between the outer β-strands of propeller blades two and

three of one monomer and the equivalent NCS-related β-strands of the other. A continuous antiparallel 8-stranded β-sheet is formed between the four β-strands of propeller three (residues 265-304), and the four β-strands of the equivalent NCS-related propeller blade (Figure 7.14). Hence, the nitrite reductase dimer can be described as having an extended β-interface (Jones and Thornton, 1995). Eight inter-monomer main-chain/main-chain hydrogen bonds are made between residues 296, 298, 300 and 302, located within the two outermost β-strands (3-4) of the propeller domains. Additional hydrogen bonds are made by the side-chains of surrounding residues (Table 7.7). These contacts are supplemented by interactions between the two short β-strands (residues 251-260) inserted between blades two and three, with the NCS-related outermost strand of blade two (2-4; residues 245-249). Interactions are also formed between the helix of the linker region (residues 105-110) and the loop between the two short inserted β-strands. Potential hydrogen bonds across the dimer interface are presented in Table 7.7.

**Table 7.7** Potential hydrogen bonds between the NIR monomers[a]

| Source Atom (Monomer A) | Location of Source Atom | Distance (Å) | Target Atom (Monomer B) | Location of Target Atom |
|---|---|---|---|---|
| 105 Leu N | Linker α-helix | 3.1 | 257 Gly O | β-strand i2:3[b] |
| 246 Lys NZ | β-strand 2-4 | 3.1 | 304 Glu OE2 | β-strand 3-4 |
| 294 Asp OD2 | β-strand 3-4 | 2.8 | 340 Lys NZ | β-strand 4-4 |
| 296 Ser O | β-strand 3-4 | 2.9 | 304 Glu N | β-strand 3-4 |
| 297 Asp OD1 | β-strand 3-4 | 2.9 | 341 Leu N | β-strand 4-4 |
| 298 Leu N | β-strand 3-4 | 2.8 | 302 Thr O | β-strand 3-4 |
| 298 Leu O | β-strand 3-4 | 2.7 | 302 Thr N | β-strand 3-4 |
| 300 Thr N | β-strand 3-4 | 3.2 | 300 Thr O | β-strand 3-4 |
| 300 Thr OG1 | β-strand 3-4 | 3.3 | 300 Thr OG1 | β-strand 3-4 |

[a]Only interactions from monomer A to monomer B are given in the table. The NCS equivalent contacts are not shown.

[b] i2:3 describe the β-strands inserted between propeller blades 2 and 3.

A total of 2190 $Å^2$ of solvent accessible surface, calculated using the programs AREAIMOL and RESAREA from the CCP4 suite (Collaborative Computational Project, 1994) and assuming a solvent probe of radius 1.4 Å, is buried in this dimer interface. This figure is similar to that for the buried surfaces of the dimer interface calculated for the *P. aeruginosa* and *T. pantotropha* enzymes, i.e. 2110 $Å^2$ and 2260 $Å^2$ respectively.


**Figure 7.14** Ribbon diagrams of two orthogonal views of the *P. stutzeri* NIR dimer interface

The figure was prepared with MOLSCRIPT (Kraulis, 1991).

Electron density maps reveal the presence of ordered water molecules in the cavity formed between the secondary structure elements of the dimer interface (see lower diagram in Figure 7.14).

All $cd_1$-type nitrite reductases isolated to date have been found to be dimers in solution. The non-covalent dimer interface is stable even under severe conditions. The *P. aeruginosa* nitrite reductase dimer prevails in 3M NaCl, 1M $Na_2SO_4$, or 6M urea,

with dissociation to monomers only occurring at very high pH (>11) or after succinylation (Kuronen et al., 1975). The suggestion has been raised that the dimeric state found for nitrite reductase is an artifact owing to its isolation in aqueous solution, and that the interface involved in forming the dimer may in fact interact with the cytoplasmic membrane *in vivo*. If this were the case then nitrite reductase may act as a monomer. Silvestrini and colleagues have prepared monomeric *P. aeruginosa* nitrite reductase which was active, indicating that dimer formation is not obligatory for enzymatic function (Silvestrini et al., 1995).

## 7.7 ASPECTS OF THE CRYSTAL STRUCTURE

### *7.7.1 Crystal contacts*

The crystal contacts in the frozen nitrite reductase lattice are principally formed between the β-propeller domain of monomer A with its symmetry-related equivalents, and between this domain and the N-terminal *c* heme-binding domain of monomer B. The β-propeller domain of monomer B, and the N-terminal domain of monomer A make no hydrogen bonding contacts in the lattice (Table 7.8). A number of residues (Section 6.7.14) have not been modelled for the structure, particularly for the N-terminal domain and it is possible that some of these missing atoms are involved in crystal contacts; if this was the case, however, it would be expected that these regions would be the most ordered and hence would be prominent in the electron density maps. It is interesting to note that two of the regions (residues 386-392 and 466-469) of the *P. stutzeri* β-propeller domain that contain insertions relative to the sequences of the *P. aeruginosa* and *T. pantotropha* enzymes are involved in crystal contacts.

The contacts formed in the room temperature cell are essentially the same as those observed for the frozen cell.

**Table 7.8** Potential hydrogen bonds between NIR dimers in the frozen and room temperature lattices

| Source Atom | Target Atom | Distance (Å) 113K | Distance (Å) Room Temp. |
|---|---|---|---|
| A113 Asn OD1 | A337 Lys NZ | 3.6 | 3.2 |
| A114 Leu O | A337 Lys NZ | 3.3 | 3.7 |
| A243 Glu OE1 | A388 Ser OG | 2.7 | 2.9 |
| A243 Glu OE1 | A388 Ser N | 3.1 | 3.3 |
| A448 Ser O | B6 Thr OG1 | 2.6 | 2.6 |
| A449 Ser OG | B9 Glu OE2 | 3.5 | 3.5 |
| A450 Ala N | B9 Glu OE2 | 2.5 | 2.5 |

### *7.7.2 Disorder in protein crystals*

Two types of disorder can exist in protein crystals: i) static disorder where the position of a domain or structural fragment changes from one unit cell to another and ii) dynamic disorder where the position of the fragment is changing in time. For X-ray diffraction experiments the measured intensities are averaged over time (i.e. the length of the data collection) and space (i.e. over all unit cells of the crystal), and hence it is not possible to differentiate these two effects. For a frozen crystal static disorder would be expected to be the predominant factor.

It is not an uncommon occurrence for multidomain proteins to have disordered domains. Some examples from the literature are the chaperonin GroEL (Braig et al., 1995), the rat phospholipase C-δ1 (Essen et al., 1996), the ribosomal protein L9 (Hoffman et al., 1996), the Sindbis virus core protein (Choi et al., 1991), the diphtheria *tox* repressor (Ding et al., 1996), the T-cell receptor - HLA-A2 - viral peptide complex (Garboczi et al., 1996) and apo-lactoferrin (B. F. Anderson, personal communication). One example which appears strikingly similar to that of the *P. stutzeri* nitrite reductase is the structure of the yeast cytochrome $b_2$. This redox protein is a tetramer, with each monomer comprised of two domains connected by a linker region. The N-terminal domain is a heme-binding cytochrome $b_5$ domain, whereas the C-terminal domain is a $\beta_8\alpha_8$ barrel that binds FMN (Xia and Mathews, 1990; Tegoni and Cambillau, 1994). Of the two monomers contained within the asymmetric unit, one of the cytochrome $b_5$-like domains is completely disordered. Interest-

ingly, if the two domains are separated by proteolytic cleavage they do not then reassociate or participate in redox reactions (Silvestrini et al., 1993).

### *7.7.3 Rationalization of the anisotropic diffraction*

From inspection of both the room temperature and frozen crystal lattices (whose dimensions are summarized in Table 7.9)

**Table 7.9** Cell dimensions for the room temperature and frozen crystals

| | a (Å) | b (Å) | c (Å) | α (°) | β (°) | γ (°) |
|---|---|---|---|---|---|---|
| Room Temperature | 80.6 | 84.6 | 105.9 | 90.0 | 98.7 | 90.0 |
| Liquid Nitrogen | 75.6 | 82.3 | 102.8 | 90.0 | 94.4 | 90.0 |

a physical interpretation of the anisotropy in the observed diffraction patterns can be presented. The β-propeller domains of the nitrite reductase dimer pack side-by-side, forming chains in the crystal running along the *c* axis with the non-crystallographic two-fold axis perpendicular to the direction of the chain. The contacts formed in this direction are between the relatively well-ordered β-propeller domains (Figure 7.15). The improvement in the diffraction along *c** as a result of freezing the crystals (from ~3.3 Å to ~2.6 Å) is associated with the shortening of the *a* and *c* cell lengths. In the room temperature cell the longer *a* axis, as well as the looser packing along *c,* results in the β-propeller domains not lining up exactly to form chains (as evident by the fact that the NCS two-fold axis is no longer perpendicular to *c**). The shift in position of the N-terminal domain of monomer B that occurs upon crystal freezing is also due to the dehydration and consequent shrinkage of the cell. The weak diffraction observed along *a** arises because the contacts formed in this orientation are only between the disordered N-terminal domains and a β-propeller domain from a symmetry related molecule.

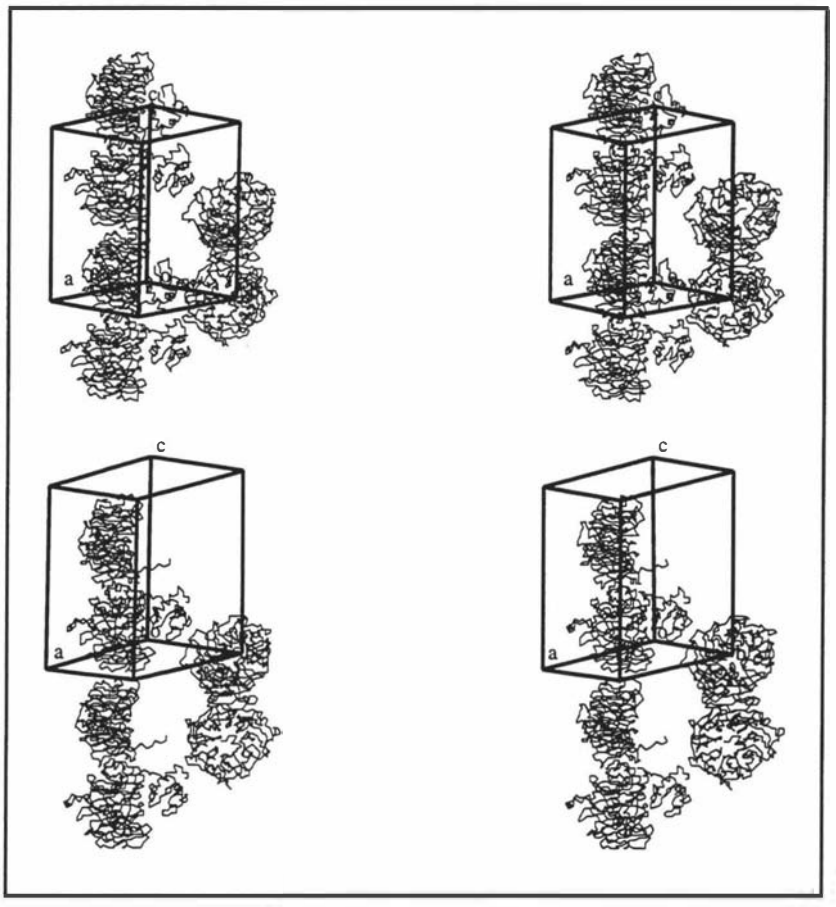



**Figure 7.15** Stereo diagrams of the frozen (upper) and room temperature (lower) NIR lattices.

This figure was prepared with TURBO-FRODO (Roussel and Cambillau, 1991).

## 7.8 FUNCTION

### *7.8.1 Flexibility of P. stutzeri nitrite reductase*

Flexibility of structure is important for the biological function of many proteins, be it binding of a substrate, binding to a receptor, or formation of a macromolecular complex. Often this flexibility is manifested by the movements of amino acid side-chains, or of large portions of the structure, either complete subunits or domains. Examples of proteins in which large conformational changes are important for biological function include hemoglobin, lactoferrin, citrate synthase, calmodulin, the immunoglobulins and the chaperonin GroEL (Huber, 1987, Gerstein et al., 1994, and the Database of Macromolecular Movements, http://bioinfo.mbb.yale.edu/MolMovDB/, M. Gerstein and W. G. Krebs, unpublished results). It is difficult, however, to ratio-

nalize the rigid-body shift observed between the *c* and $d_1$ domains for the *P. stutzeri* nitrite reductase in terms of a reaction mechanism involving electron transfer between the hemes. A possibility to consider is that the binding of the external electron donor, cytochrome $c_{551}$ or azurin, to the N-terminal domain may result in the movement of this domain, packing it against the β-propeller domain and bringing the *c* heme and $d_1$ heme into a suitable position and orientation for electron transfer.

### *7.8.2 Electron transfer between the heme groups*

The distances between the heme groups of the dimer are consistent with the path of electron transfer being between the *c* heme and the $d_1$ heme of the same monomer. This is in agreement with the observation that a preparation of monomeric *P. aeruginosa* nitrite reductase retains the catalytic and spectral properties exhibited by the dimer (Silvestrini et al., 1995). Allosteric behaviour for the enzyme from *P. aeruginosa* has been observed on carbon monoxide binding (Parr et al., 1975), as well as when oxidised with cytochrome $c_{551}$ or azurin (Blatt and Pecht, 1979). This behaviour has been ascribed to the presence of the N-terminal arm that is swapped between the monomers of this enzyme (Nurizzo et al., 1997). No evidence of any allosteric behaviour has been reported for the nitrite reductase from any other species, though no detailed stop flowed studies have been conducted on these enzymes.

The rate of electron transfer between protein redox centres has been shown to be related to the separation and orientation of the centres (Makinen et al., 1983; Beratan et al., 1992; Moser et al., 1992). Cofactors in redox complexes, either contained within the same protein or within macromolecular complexes, appear to be grouped into two predominant orientations. The cytochrome *c* peroxidase - cytochrome *c* complex (Pelletier and Kraut, 1992), the heme and flavin containing *p*-cresol methylhydroxylase (Mathews et al., 1991), and the nitrite reductases from *P. aeruginosa* and *T. pantotropha* have their respective redox cofactors arranged at an angle of 60-70°. The flavocytochrome $b_2$ (Xia and Mathews, 1990) and dimeric cytochrome *c'* (Weber et al., 1980) have the redox centres in a co-planar arrangement. The internal electron transfer from the *c* to $d_1$ heme of *P. aeruginosa* nitrite reductase is unusually slow (Parr et al., 1977; Silvestrini et al., 1982; Silvestrini et al., 1990; Schichman et al., 1996). A recent study of the enzyme from *T. pantotropha* has revealed a faster rate (Kobayashi et al., 1997), though in the light of the crystal structures of these two enzymes (Fulop et al., 1995; Nurizzo et al., 1997) it might be the

case that these measurements have been conducted on different initial starting structures. The cytochrome *c* oxidases from bacteria (Ostermeier et al., 1997) and mitochondria (Tsukihara et al., 1996) have the *a* heme cofactors at an angle of ~105°, though the redox behaviour of these enzymes is complicated by the additional copper redox centres. The $d_1$ heme - *c* heme orientations of the *P. stutzeri* nitrite reductase appear to fall outside these general classifications, and the distances between the two heme groups are much larger than that found for the NIRs from other species (Table 7.10).

**Table 7.10** Relationship of the *c* and $d_1$ heme groups within the NIR monomers

| NIR model | $d_1$ heme - *c* heme angle (°) | Fe...Fe distance (Å) | $d_1$ - *c* heme edge distance (Å) |
|---|---|---|---|
| *T. pantotropha* | 62 | 20.6 | 9.8 |
| *P. aeruginosa* | 70 | 19.6 | 11.4 |
| *P. stutzeri* - monomer A | 55 | 24.1 | 19.4 |
| *P. stutzeri* - monomer B | 27 | 26.7 | 22.2 |

The nature of the polypeptide that separates the two redox centres between which electron transfer occurs has been the subject of much interest. It is unclear as to whether the electron is transferred in a through-bonds mechanism with the polypeptide bonded orbitals acting as a wire (Beratan et al., 1992), or whether a through-space mechanism exists. It has been tentatively proposed that the non-planar arrangement of the redox centres may result in an increased ability to store electrons prior to a reduction step (Makinen et al., 1983), and that the slow electron transfer rate, in conjunction with inhibition by NO, may be an adaptation to prevent increased concentrations of this toxic product in the cell. The reaction product NO is able to rebind to the $d_1$ and *c* hemes under certain conditions (Silvestrini et al., 1979). Hence the catalytic activity of nitric oxide reductase will most likely influence the reaction rate of nitrite reductase. It may be the case that during bacterial denitrification one of the other enzymes in this pathway may catalyse the rate determining step for the total reduction of nitrate to nitrogen gas, and hence no advantage is gained by selecting for a faster nitrite reductase.

The differences in the relative positions and orientations of the *c* hemes of the two monomers of *P. stutzeri* nitrite reductase when compared to their $d_1$ hemes would result in asymmetry in the rate of electron transfer in the two subunits of the dimer if this conformation is retained in solution. It is more likely the case that the conforma-

tions observed in this crystal structure are a result of crystal packing and do not represent a functional state. However, for the molecules to pack in this way there must exist inherent flexibility in the linker region that connects the two domains. It may the case that the *P. stutzeri* NIR does not adopt the conformation required for electron transfer until the cytochrome $c_{551}$ or azurin electron donor has bound.

The position of the N-terminal domain found for the *P. aeruginosa* and *T. pantotropha* enzymes would introduce steric clashes with symmetry related β-propeller domains if this was the orientation adopted for the N-terminal domains of the *P. stutzeri* enzyme. These clashes are not severe and involve only a few residues (residues 59-61) of each domain. These residues do, however, precede the large twelve-residue insertion which could be affecting the packing of the molecules in the lattice.

### *7.8.3 Absence of an N-terminal extension*

The N-terminal regions of the bacterial nitrite reductases are the least well conserved regions of the enzyme. From the crystal structures determined to date for the two nitrite reductases that contain this extension, it appears that this part of the structure is critical for enzymatic function. In the *T. pantotropha* oxidised enzyme the N-terminus provides a tyrosine (residue 25) that is bound to the $d_1$ heme iron and plays a crucial role in the proposed reaction mechanism (Fulop et al., 1995). In the reduced structure of the enzyme, however, the N-terminal region is disordered and the tyrosine is no longer coordinated (Williams et al., 1997). The N-terminal extension in *P. aeruginosa* nitrite reductase does not provide a $d_1$ heme ligand directly, though a homologous tyrosine (residue 10) does hydrogen bond to the water molecule/hydroxide ion that is bound to the $d_1$ heme iron. The N-terminal arm for this enzyme is 'domain swapped' to the other subunit of the monomer (Nurizzo et al., 1997). As well as appearing to provide specific interactions to and around the $d_1$ heme, the N-terminal extension also serves to limit the solvent accessibility of the heme group in the oxidised enzymes from these species.

The deletion of this N-terminal region from the *P. stutzeri* nitrite reductase raises many questions as to how the functional properties of the enzyme are modified. The absence of an N-terminal tyrosine that could interact with the $d_1$ heme had suggested that a somewhat different mechanism existed for this enzyme. However, a recent site-directed mutagenesis experiment, in which the N-terminal tyrosine (residue 10) was mutated to a phenylalanine in the *P. aeruginosa* enzyme, indicated that this amino acid was not crucial for function, with the mutant enzyme showing similar

activity and spectral properties to that of the wild-type protein (Cutruzzola et al., 1997). This result, along with the absence of a homologous residue in the *P. stutzeri* and *A. eutrophus* enzymes, indicates that *T. pantotropha* nitrite reductase appears to be unique as regards the importance of this residue. It is intriguing to note, however, that *T. pantotropha* is the only bacterial species reported to date that undergoes aerobic denitrification (Robertson and Kuenen, 1984), suggesting that the tyrosine ligand may play a role in substrate specificity, in order to limit the competing oxygen reduction that nitrite reductase is also able to catalyse.

The N-terminal domain of the *T. pantotropha* nitrite reductase also provides one of the histidine ligands to the *c* heme in the oxidised state (His 17), resulting in the unusual bis-histidine ligation of the heme. Upon reduction, the N-terminal extension, and parts of the *c* heme-binding domain, undergo a structural rearrangement in which His 17 is replaced by Met 106 as the *c* heme iron ligand, and Tyr 25 is removed from coordination to the $d_1$ heme ligand (Williams et al., 1997). The enzyme from *P. aeruginosa* has a histidine (residue 21) in a similar position on the N-terminal extension. This residue does not coordinate the *c* heme in the oxidised state. Clearly an identical structural rearrangement cannot occur as the *c* heme iron is already coordinated by an equivalent methionine (residue 88). In the absence of any structural information on the reduced state it is unclear whether the tyrosine moves away from the $d_1$ heme, which in this case is actually that of the other monomer in the dimer. The lack of the N-terminal extension in *P. stutzeri* nitrite reductase clearly removes any possibility of an alternative *c* heme ligand coming from this part of the structure. The effect of this deletion on the structure of the reduced enzyme is unknown. Whether there is some sort of rearrangement in the ferrous state, as found for the *T. pantotropha* enzyme, or whether the fold of the domain remains essentially unchanged upon reduction, as is found for the class I *c*-type cytochromes (Matsuura et al., 1982; Takano and Dickerson, 1981), must await the determination of the structure of the reduced enzyme from *P. stutzeri.*

From sequence considerations it is tempting to suggest that the twelve-residue insertion (residues 45-56) in the N-terminal domain of *P. stutzeri* NIR may serve to close off the $d_1$ heme in a similar manner to the N-terminal extension of the *P. aeruginosa* and *T. pantotropha* enzymes. The rigid-body shift of the N-terminal domain does in fact place this region of the structure in a position where this could be possible, though apart from the large unexplained density in the region of the $d_1$ heme there is little crystallographic evidence to support this. For the *P. aeruginosa* and *T.*

*pantotropha* enzymes the homologous region of the N-terminal domain from which this insertion originates is located towards the dimer interface, well removed from the $d_1$ heme pocket.

## 7.9 SUGGESTIONS FOR FURTHER STUDIES

The immediate priority must be to finish the structure of the oxidised *P. stutzeri* nitrite reductase and obtain a complete model of the N-terminal domain. It would be most desirable, and would make this objective more easily attainable, if an improved native data set could be obtained. In order to alleviate the problem of N-terminal domain disorder in the crystal lattice it may be necessary to proteolytically cleave the two domains that form the monomer, purify and then attempt to crystallize just the N-terminal domain. Determining the structure of the *c* heme-binding domain in this manner would, of course, result in the loss of any information regarding $d_1$ heme - *c* heme distances and orientations, important for understanding the biological function of the enzyme as regards electron transfer, but it may then prove possible to rigid-body refine the completed structure of the domain into the native whole molecule cell. If crystallization experiments proved unsuccessful, a cleaved N-terminal domain would be small enough to be suitable for an NMR structure determination, as has been successfully done for the structures of cytochrome $c_{551}$ from *P. stutzeri* (Cai et al., 1992) and *P. aeruginosa* (Timkovich and Cai, 1993).

With a complete model of the oxidised nitrite reductase structure it would then be possible to investigate the interactions between the physiological electron donors, cytochrome $c_{551}$ and pseudoazurin. The nature of these interactions are of interest because it is assumed that the binding of the electron donor must occur in such a way as to orient the electron-donating redox centre for electron transfer. However the fact that at least two classes of proteins, the *c*-type cytochromes and azurins, with different protein folds and types of redox centre can act in this role, suggest that the binding site may be 'promiscuous' in nature. The structure of cytochrome $c_{551}$ from *P. stutzeri* has been determined (Cai et al., 1992), so docking studies could be conducted to model complex formation between this molecule and the nitrite reductase, in an effort to understand the mechanism of electron transfer between these two proteins. It would be very interesting to attempt the crystallization of the cytochrome $c_{551}$ - nitrite reductase redox complex. This method has been successful in the determination of the structure of other redox complexes, for example: the yeast cytochrome *c* - cytochrome *c* peroxidase complex (Pelletier and Kraut, 1992) and

the amicyanin - cytochrome $c_{551}$ - methylamine dehydrogenase complex from *Paracoccus denitrificans* (Chen et al., 1994). A cautionary note is that, not surprisingly, it appears important that both the redox proteins are from the same species to form a stable complex. Co-crystallization experiments of yeast cytochrome *c* peroxidase and horse heart cytochrome *c* yielded a lattice containing both redox proteins, but with the cytochrome *c* molecules disordered (Poulos et al., 1987; Pelletier and Kraut, 1992). As the cytochrome $c_{551}$ interacts with the *c* heme-binding domain of nitrite reductase the formation of a complex between these molecules may reduce the disorder exhibited by this domain in the native oxidized structure.

A structural characterization of the reduced *P. stutzeri* nitrite reductase is essential for a complete understanding of the function of this enzyme. It might prove possible to reduce the nitrite reductase in the oxidised crystal form. If any structural rearrangement of the N-terminal domain does occur, it might do so without affecting the integrity of the crystal lattice as the contacts are formed predominantly by the β-propeller domains. The variable stability of the crystals, as observed by their behaviour once removed from the crystallization drop, may make these experiments difficult. It may prove more desirable to reduce the protein prior to crystallization experiments in the hope of obtaining a different crystal form that is more amenable to X-ray crystallographic studies.

The binding of nitrite (or substrate mimics), to the $d_1$ heme of the reduced nitrite reductase, and subsequent trapping of the nitric oxide reaction product, could be attempted by freezing the protein crystals soaked in substrate to quench the reaction in a manner analogous to the experiments conducted for the *T. pantotropha* enzyme (Williams et al., 1997). These kinds of experiments would provide valuable information regarding the reaction mechanism of the enzyme.

Crystallization of the nitrite reductase from the closely related *P. stutzeri* sub-species, *P. stutzeri* ZoBell, may result in the formation of crystals that are more suited for X-ray diffraction experiments. The two proteins are similar enough that structural information obtained for one enzyme could be utilised in the determination of the structure of the other. Any functional conclusions drawn from one of these structures would be applicable to both enzymes.

It is important that the gene encoding the *P. stutzeri* JM300 nitrite reductase be resequenced in order to resolve the conflicts between the published sequence and the

X-ray structure. Some of these observed differences involve residues thought to play a key role in the mechanism of this enzyme. In order to understand, and form conclusions about, the functions of this protein and make comparisons to homologous enzymes it is essential to know the identity of the amino acids residues present.

Owing to the appearance of extra electron density for the $d_1$ heme in the electron density maps further characterization of this prosthetic group is necessary. Further mass spectroscopy experiments are necessary to obtain the molecular weight of the $d_1$ heme and determine whether the extra density observed in the Fourier maps originates from the heme group, or from polypeptide (or some other molecule) in close proximity. A full structural characterization of the isolated $d_1$ heme by X-ray crystallographic and/or NMR methods (in the manner of Barkigia et al., 1992) may be necessary. It is of interest to determine if the heme is a novel protein cofactor and whether it has been modified to 'replace' the missing tyrosine interaction with the heme iron.

The *P. aeruginosa* nitrite reductase has long been regarded as the archetypal nitrite reductase with much biochemical and functional data measured for this enzyme. However, with the recent structural characterization of three of these enzymes, it appears that substantial differences exist between the members of this protein family. Further biochemical measurements for *P. stutzeri* nitrite reductase are required if the structural differences observed are to be correlated with alterations in protein function.

## REFERENCES


Adair, G. S. (1925). The osmotic pressure of haemoglobin in the absence of salts. *Proceedings of the Royal Society, London.*, 109A:292-300.

Adams, P. D., Pannu, N. S., Read, R. J., and Brunger, A. T. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proceedings of the National Academy of Sciences, USA*, 94:5018–5023.

Alefounder, P. R. and Ferguson, S. J. (1980). The location of dissimilatory nitrite reductase and the control of dissimilatory nitrite reductase by oxygen in *Paracoccus denitrificans*. *Biochemical Journal*, 192:231–240.

Amiconi, G., Ascoli, F., Barra, D., Bertollini, A., Matarese, R. M., Verzili, D., and Brunori, M. (1989). Selective oxidation of methionine β(55)D6 at the $\alpha_1\beta_1$ interface in hemoglobin completely destabilizes the T-state. *The Journal of Biological Chemistry*, 264(30):17745–17749.

Antonini, E., Rossi-Bernardi, L., and Chiancone, E., editors (1981). *Hemoglobins*, volume 76 of *Methods in Enzymology*. Academic Press.

Arai, H., Igarashi, Y., and Kodama, T. (1991a). Anaerobically induced expression of the nitrite reductase cytochrome *c*-551 operon from *Pseudomonas aeruginosa*. *FEBS Letters*, 280(2):351–353.

Arai, H., Igarashi, Y., and Kodama, T. (1991b). Nitrite activates the transcription of the *Pseudomonas aeruginosa* nitrite reductase and cytochrome *c*-551 operon under anaerobic conditions. *FEBS Letters*, 288(1):227–228.

Amone, A. (1972). X-ray diffraction study of binding of 2,3-diphosphoglycerate to human deoxyhaemoglobin. *Nature*, 237:146–149.

Aronson, H-E. G., Royer, W. E., and Hendrickson, W. A. (1994). Quantification of tertiary structural conservation despite primary sequence drift in the globin fold. *Protein Science*, 3:1706–1711.

Averill, B. A. (1996). Dissimilatory nitrite and nitric oxide reductases. *Chemical Reviews*, 96:2951–2964.

Baker, S. C., Saunders, N. F. W., Wills, A. C., Ferguson, S. J., Hajdu, J., and Fulop, V. (1997). Cytochrome $cd_1$ structure: unusual haem environments in a nitrite reductase and analysis of factors contributing to β-propeller folds. *Journal of Molecular Biology*, 269:440–455.

Baldwin, J. and Chothia, C. (1979). Haemoglobin: the structural changes related to ligand binding and its allosteric mechanism. *Journal of Molecular Biology*, 129:175–220.

Baldwin, J. M. (1980). The structure of human carbonmonoxy haemoglobin at 2.7 Å resolution. *Journal of Molecular Biology*, 136:103–128.

Baralle, F. E., Shoulders, C. C., and Proudfoot, N. J. (1980). The primary structure of the human ε-globin gene. *Cell*, 21:621–626.

Barber, D., Parr, S. R., and Greenwood, C. (1976). Some spectral and steady-state kinetic properties of *Pseudomonas* cytochrome oxidase. *Biochemical Journal*, 157:431–439.

Barcroft, J. (1928). *The Respiratory Function of the Blood.*, volume Part II. London, Cambridge University Press.

Barkigia, K. M., Chang, C. K., Fajer, J., and Renner, M. W. (1992). Models of heme $d_1$. Molecular structure and NMR characterization of an iron (iii) dioxoisobacteriochlorin (porphyrindione). *Journal of the American Chemical Society*, 114:1701–1707.

Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Engineering*, 6(1):37–40.

Bauer, C., Ludwig, I., and Ludwig, M. (1968). Different effects of 2,3-diphosphoglycerate and adenosine triphosphate on the oxygen affinity of adult and foetal human haemoglobin. *Life Sciences*, 7:1339-1343.

Bauer, C., Tamm, R., Petschow, D., Bartels, R., and Bartels, H. (1975). Oxygen affinity and allosteric effects of embryonic mouse haemoglobins. *Nature*, 257:333–334.

Benesch, R., Benesch, R. E., and Enoki, Y. (1968a). The interaction of hemoglobin and its subunits with 2,3-diphosphoglycerate. *Proceedings of the National Academy of Science, USA*, 61:1102-1106.

Benesch, R., Benesch, R. E., and Yu, C. I. (1968b). Reciprocal binding of oxygen and diphosphoglycerate by human hemoglobin. *Proceedings of the National Academy of Science, USA*, 59:526-532.

Benesch, R. E., Benesch, R., Renthal, R. D., and Maeda., N. (1972). Affinity labeling of the polyphosphate binding site of hemoglobin. *Biochemistry*, 11:3576-3582.

Benesch, R. E., Benesch, R., and Yu, C. I. (1969). The oxygenation of hemoglobin in the presence of diphosphoglycerate. Effect of temperature, pH, ionic strength and hemoglobin concentration. *Biochemistry*, 8:2567-2571.

Beratan, D. N., Onuchic, J. N., Winkler, J. R., and Gray, H. B. (1992). Electron-tunneling pathways in proteins. *Science*, 258:1740–1741.

Berger, H. and Wharton, D. C. (1980). Small-angle X-ray scattering studies of oxidised and reduced cytochrome oxidase from *Pseudomonas aeruginosa*. *Biochimica et Biophysica Acta.*, 622:355–359.

Berks, B. C., Ferguson, S. J., Moir, J. W. B., and Richardson, D. J. (1995a). Enzymes and associated electron transport systems that catalyse the respiratory reduction of nitrogen oxides and oxyanions. *Biochimica et Biophysica Acta*, 1232:97–173.

Berks, B. C., Page, M. D., Richardson, D. J., Reilly, A., Cavill, A., Outen, F., and Ferguson, S. J. (1995b). Sequence analysis of subunits of the membrane-bound nitrate reductase from a denitrifying bacterium: the integral membrane subunit provides a prototype for the dihaem electron carrying arm of a redox loop. *Molecular Microbiology*, 15:319–331.

Blatt, Y. and Pecht, I. (1979). Allosteric cooperative interactions among redox sites of *Pseudomonas* cytochrome oxidase. *Biochemistry*, 18(13):2917–2922.

Blessing, R. H., Guo, D. Y., and Langs, D. A. (1996). Statistical expectation value of the Debye-Waller factor and E(hkl) values for macromolecular crystals. *Acta Crystallographica*, D52:257–266.

Blessing, R. H. and Langs, D. A. (1988). A priori estimation of scale and overall anisotropic temperature factors from the Patterson origin peak. *Acta crystallographica*, A44:729–735.

Bohr, C. (1904). Theoretische behandlung der quantitativen verhaltnis bei der sauerstoffaufnahme des hamoglobins. *Zbl. Physiol.*, 17:682.

Bohr, C., Hasselbalch, K., and Krogh, A. (1904). Ueber einien in biologischer beziehung wichtigen einfluss, den die kohensaurespannung des blutes auf dessen sauerstoffbindung ubt. *Skand. Arch. Physiol.*, 16:402.

Bonaventura, C., Arumugam, M., Cashon, R., Bonaventura, J., and Moo-Penn, W. F. (1994). Chloride masks effects of opposing positive charges in Hb A and Hb Hinsdale (β139 Asn→Lys) that can modulate cooperativity as well as oxygen affinity. *Journal of Molecular Biology*, 239:561–568.

Bonaventura, C. and Bonaventura, J. (1978). Anionic control of hemoglobin function. In Caughy, W. S., editor, *Biochemical and Clinical Aspects of Hemoglobin Abnormalities*, pages 647–663. Academic Press, New York.

Braig, K., Adams, P. D., and Brunger, A. T. (1995). Conformational variability in the refined crystal structure of the chaperonin GroEL at 2.8 Å resolution. *Nature Structural Biology*, 2:1083–1094.

Branden, C. I. and Jones, T. A. (1990). Between objectivity and subjectivity. *Nature*, 343:687–689.

Braunitzer, G. (1958). The primary structure of the protein components of several haemoglobins. *Z. Physiol. Chemie*, 312:72-84.

Bricogne, G. (1993). Direct phase determination by entropy maximization and likelihood ranking: status report and perspectives. *Acta Crystallographica*, D49:37–60.

Brittain, T., Blackmore, R., Greenwood, C., and Thomson, A. J. (1992). Bacterial nitrite-reducing enzymes. *European Journal of Biochemistry*, 209:793–802.

Brittain, T., Hofmann, O. M., Watmough, N. J., Greenwood, C., and Weber, R. E. (1997). A two-state analysis of co-operative oxygen binding in the three human embryonic haemoglobins. *Biochemical Journal*, 326:299–303.

Brittain, T., Sutherland, J., and Greenwood, C. (1986). A study of the kinetics of the reaction of ligands with the liganded states of mouse embryonic haemoglobins. *Biochemical Journal*, 234:151–155.

Brock, T. D., Madigan, M. T., Martinko, J. M., and Parker, J. (1997). *Biology of Microorganisms.* Prentice-Hall, Inc.

Brown, J. L. and Ingram, V. M. (1974). Structural studies on chick embryonic hemoglobins. *The Journal of Biological Chemistry*, 249:3960–3972.

Brunger, A. T. (1990). Extension of molecular replacement: a new search strategy based on Patterson correlation refinement. *Acta Crystallographica*, A46:46–57.

Brunger, A. T. (1992a). Free R-value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355:472–475.

Brunger, A. T. (1992b). *X-PLOR Version 3.1. A system for X-ray crystallography and NMR.* Yale University Press, U.S.A.

Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R., .Jiang, J-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. (1997). Crystallography and NMR system (CNS).

Brunger, A. T., Kuriyan, J., and Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science*, 235:458–460.

Brunori, M., Antonini, G., Malatesta, F., Sarti, P., and Wilson, M. T. (1988). Structure and function of cytochrome oxidase: a second look. In Eichorn, G. L. and Marzilli, L. G., editors, *Heme Proteins.* Elsevier Science Publishing.

Bunn, H. F. and Briehl, R. W. (1970). The interaction of 2,3-diphosphoglycerate with various human hemoglobins. *Journal of Clinical Investigation*, 49:1088-1095.

Bunn, H. F., Forget, B. G., and Ranney, H. M. (1977). *Human Hemoglobins*. W. B. Saunders Company.

Cai, M., Bradford, E. G., and Timkovich, R. (1992). Investigation of the solution conformation of cytochrome *c*-551 from *Pseudomonas stutzeri*. *Biochemistry*, 31:8603–8612.

Capp, G. L., Rigas, D. A., and Jones, R. T. (1967). Hemoglobin Portland 1: a new hemoglobin unique in structure. *Science*, 157:65–66.

Capp, G. L., Rigas, D. A., and Jones, R. T. (1970). Evidence for a new haemoglobin chain. *Nature*, 228:278–279.

Chang, C. K., Timkovich, R., and Wu, W. (1986). Evidence that heme $d_1$ is a 1,3-porphyrindione. *Biochemistry*, 25:8447–8453.

Chapman, B. S., Tobin, A. J., and Hood, L. E. (1980). Complete amino acid sequences of the major early embryonic α-like globins of the chicken. *The Journal of Biological Chemistry*, 255:9051–9059.

Cheesman, M. R., Ferguson, S. J., Moir, J. W. B., Richardson, D. J., Zumft, W. G., and Thomson, A. J. (1997). Two enzymes with a common function but different heme ligands in the forms as isolated. Optical and magnetic properties of the heme groups in the oxidised forms of nitrite reductase, cytochrome $cd_1$, from *Pseudomonas stutzeri* and *Thiosphaera pantotropha*. *Biochemistry*, 36:16267–16276.

Chen, L., Durley, R. C. E., Mathews, F. S., and Davidson, V. L. (1994). Structure of an electron transfer complex: methylamine dehydrogenase, amicyanin and cytochrome $c_{551i}$. *Science*, 264:86–90.

Chernoff, A. I. (1961). The amino acid composition of hemoglobin. An improved method for separating the peptide chains of human hemoglobin. *Journal of Chromatography*, 6:252-257.

Choi, H-K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M. G., and Wengler, G. (1991). Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. *Nature*, 354:37–43.

Clegg, J. B. and Gagnon, J. (1981). Structure of the ζ chain of human embryonic hemoglobin. *Proceedings of the National Academy of Science, USA*, 78(10):6076–6080.

Clegg, J. B., Naughton, M. A., and Weatherall., D. J. (1966). Abnormal human hemoglobins. Separation and characterization of the α and β chains by chromatography, and the determination of two new variants Hb Chesapeake and Hb J (Bangkok). *Journal of Molecular Biology*, 19:91-108.

Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica*, D50:760–763.

Cosier, J. and Glazer, A. M. (1986). A nitrogen-gas-stream cryostat for general X-ray diffraction studies. *Journal of Applied Crystallography*, 19:105–107.

Cowtan, K. D. (1996). Review: cross-validation and the free-R. In Dodson, E., Moore, M., Ralph, A., and Bailey, S., editors, *Proceedings of the CCP4 Study Weekend*, pages 23–28. CCLRC Daresbury Laboratory, Warrington, U.K.

Cowtan, K. D. and Main, P. (1996). Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallographica*, D52:43–48.

Crane, B. R., Arvai, A. S., Ghosh, D. K., Wu, C., Getzoff, E. D., Stuehr, D., and Tainer, J. A. (1998). Structure of nitric oxide synthase oxygenase dimer with pterin and substrate. *Science*, 279:2005–2168.

Crennell, S. J., Garmen, E. F., Laver, W. G., Vimr, E. R., and Taylor, G. L. (1993). Crystal structure of a bacterial sialidase (from *Salmonella typhimurium* LT2) shows the same fold as an influenza neuraminidase. *Proceedings of the National Academy of Sciences, USA*, 90:9852–9856.

Criddle, C. S., DeWitt, J. T., Grbic-Galic, D., and McCarty, P. L. (1990). Transformation of carbon tetrachloride by *Pseudomonas* sp. strain KC under denitrification conditions. *Applied and Environmental Microbiology*, 56:387–395.

Crowther, R. A. (1972). In Rossman, M. G., editor, *The Molecular Replacement Method*, pages 173–178. New York: Gordon &Breach.

Crowther, R. A. and Blow, D. M. (1967). A method of positioning a known molecule in an unknown crystal structure. *Acta Cryst.*, 23:544–548.

Cusanovich, M. A., Maeyer, T. E., and Tollin, G. (1988). *c*-type cytochromes: oxidation-reduction properties. In Eichorn, G. L. and Marzilli, L. G., editors, *Heme Proteins*. Elsevier Science Publishing.

Cutruzzola, F., Arese, M., Grasso, S., Bellelli, A., and Brunori, M. (1997). Mutagenesis of nitrite reductase from *Pseudomonas aeruginosa*: tyrosine-10 in the *c* heme domain is not involved in catalysis. *FEBS letters*, 412:365–369.

Cuypers, H. and Zumft, W. G. (1993). Anaerobic control of denitrification in *Pseudomonas stutzeri* escapes mutagenesis of an *fnr*-like gene. *Journal of Bacteriology*, 175(22):7236–7246.

Dawson, J. H. (1988). Probing structure-function relations in heme-containing oxygenases and peroxidases. *Science*, 240:433–439.

Day, C. L. (1993). *Expression and characterisation of the N-terminal half of human lactoferrin*. PhD thesis, Massey University, New Zealand.

Day, M. (1998). Superbugs take hold. *New Scientist*, 2131:4.

de Boer, A P. N., Reijnders, W. N. M., Kuenen, J. G., Stouthamer, A. H., and van Spanning, R. J. M. (1994). Isolation, sequencing and mutational analysis of a gene cluster involved in nitrite reduction in *Paracoccus denitrificans*. *Antonie van Leeuwenhoek*, 66:111–127.

Degtyarenko, D. N., North, A. C. T., Perkins, D. N., and Findlay, J. B. C. (1998). PROMISE: a database of information on prosthetic centres and metal ions in protein active sites. *Nucleic Acids Research*, 26:376–381.

Dickerson, R. E. and Geiss, I. (1983). *Hemoglobin: Structure, Function, Evolution, and Pathology*. The Benjamin/Cummings Publishing Company.

Diederichs, K. and Karplus, P. A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature Structural Biology*, 4:269–275.

Ding, X., Zeng, H., Schiering, N., Ringe, D., and Murphy, J. R. (1996). Identification of the primary metal ion-activation sites of the diphtheria *tox* repressor by X-ray crystallography and site directed mutational analysis. *Nature Structural Biology*, 3:382–387.

Di Iorio, E. E. D. (1981). Preparation of derivatives of ferrous and ferric hemoglobin. *Methods in Enzymology*, 76:57–72.

Dodson, E., Kleywegt, G. J., and Wilson, K. (1996). Report of a workshop on the use of statistical validators in protein X-ray crystallography. *Acta Crystallographica*, D52:228–234.

Drabkin, D. L. (1949). Aspects of the oxygenation and oxidation functions. In Roughton, F. J. W. and Kendrew, J. C., editors, *Haemoglobin (Barcroft Symposium)*, page 35. Interscience, New York.

Drenth, J. (1994). *Principles of Protein X-ray crystallography*. Springer-Verlag New York.

Drescher, H. and Kunzer, W. (1954). The blood pigment of the humna fetus. *Klin. Wochschr*, 32:92.

Driessen, H. P. C. and Tickle, I. J. (1994). The use of normalised amplitudes in the rotation function. In *CCP4 Newsletter May 1994*, pages 13–19. SERC Daresbury Laboratory, Warrington, U.K.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. W., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., and Proudfoot, N. J. (1980). The structure and evolution of the human β-globin family. *Cell*, 21:653–668.

Engh, R. A. and Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallographica*, A47:392–400.

Englehardt, J. F. (1925). PhD thesis. Gottingen, Germany.

Essen, L-O., Perisic, O., Cheung, R., Katan, M., and Williams, R. L. (1996). Crystal structure of a mammalian phosphoinositide-specific phopholipase Cδ. *Nature*, 380:595–602.

Faber, H. R., Groom, C. R., Baker, H. M., Morgan, W. T., Smith, A., and Baker, E. N. (1995). 1.8 Å structure of the C-terminal domain of rabbit serum hemopexin. *Structure*, 3:551–559.

Ferguson, J. K. W. and Roughton, F. J. W. (1934). The chemical relationships with physiological importance of carbamino compounds of $CO_2$ with haemoglobin. *Journal of Physiology*, 83:68-86.

Ferguson, S. (1992). The periplasm. In Mohan, S., Dow, C., and Coles, J., editors, *Prokaryotic Structure and Function*, pages 315–339. Cambridge University Press.

Fermi, G. (1975). Three-dimensional Fourier synthesis of human deoxyhaemoglobin at 2.5 Å resolution: refinement of the atomic model. *Journal of Molecular Biology*, 97:237–256.

Fermi, G., Perutz, M. F., Shaanan, B., and Fourme, R. (1984). The crystal structure of human deoxyhaemoglobin at 1.74 Å. *Journal of Molecular Biology*, 175:159–174.

Firestone, M. K. (1982). Biological denitrification. In Stevenson, F. J., editor, *Nitrogen in agricultural soils. Agronomy monograph 22*, pages 289–326. American Society for Agronomy, Madison, Wisconsin, U.S.A.

Fischer, H. and Zeile, K. (1929). Syntheses of hematoporphyrin, protoporphyrin and hemin. *Justus Leibig's Annalen der Chemie*, 468:98-116.

Fox, G. and Holmes, K. (1966). An alternative method of solving the layer scaling equations of Hamilton, Rollet and Sparks. *Acta Crystallographica*, 20:886--891.

Frausto da Silva, J. J. R. and Williams, R. J. P. (1991). *The biological chemistry of the elements: the inorganic chemistry of life*. Oxford University Press, Oxford.

French, S. and Wilson, K. (1978). On the treatment of negative intensity observations. *Acta Crystallographica*, A34:517–525.

Frier, J. A. and Perutz, M. F. (1977). Structure of human foetal deoxyhaemoglobin. *Journal of Molecular Biology*, 112:97–112.

Frolow, F., Kalb, A. J., and Yariv, J. (1994). Structure of a unique twofold symmetric haem-binding site. *Nature Structural Biology*, 270:453–460.

Fujinaga, M. and Read, R. J. (1987). Experiences with a new translation function program. *Journal of Applied Crystallography*, 20:517–521.

Fulop, V., Moir, J. W., Ferguson, S. J., and Hajdu, J. (1995). The anatomy of a bifunctional enzyme: structural basis for reduction of oxygen to water and synthesis of nitric oxide by cytochrome $cd_1$. *Cell*, 81:369–377.

Furchgott, R. F. and Zawadzki, J. V. (1980). The obligatory role of endothelial cells in the relaxation of arterial smooth muscle by acetylcholine. *Nature*, 288:373–376.

Gale, R. E., Clegg, J. B., and Huehns, E. R. (1979). Human embryonic haemoglobins Gower 1 and Gower 2. *Nature*, 280:162–164.

Garber, E. A. E. and Hollocher, T. C. (1982a). $^{15}N$,$^{18}O$ tracer studies on the activation of nitrite by denitrifying bacteria. Nitrite/water-oxygen exchange and nitrosation reactions as indicators of electrophilic catalysis. *The Journal of Biological Chemistry*, 257:8091–8097.

Garber, E. A. E. and Hollocher, T. C. (1982b). Positional isotropic equivalence of nitrogen in $N_2O$ produced by the denitrifying bacterium *Pseudomonas stutzeri*. *The Journal of Biological Chemistry*, 257:4705–4708.

Garboczi, D. N., Ghosh, P., Utz, U., Fan, Q. R., Biddison, W. E., and Wiley, D. C. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature*, 384:134–141.

Garcia, K. C., Degano, M., Pease, L. R., Hunag, M., Peterson, P. A., Teyton, L., and Wilson, I. A. (1998). Structural basis of plasticity in T cell receptor recognition of a self peptide - MHC antigen. *Science*, 279:1166–116.

Garman, E. F. and Schneider, T. R. (1997). Macromolecular cryocrystallography. *Journal of Applied Crystallography*, 30:211–237.

Gayon, U. and Dupetit, G. (1882). *Comptes Rendus de l'Academie des Sciences*, 95:644–646.

Gerstein, M., Lesk, A. M., and Chothia, C. (1994). Structural mechanisms for domain movements in proteins. *Biochemistry*, 33:6739–6749.

Ghosh, M., Anthony, C., Harlos, K., Goodwin, M. G., and Blake, C. (1996). The refined structure of the quino-protein methanol dehydrogenase from methylobacterium *extorquens* at 1.94 Å. *Structure*, 3:177–187.

Godden, J. W., Turley, S., Teller, D. C., Adman, E. T., Liu, M. Y., Payne, W. J., and LeGall, J. (1991). The 2.3 angstrom X-ray structure of nitrite reductase from *Achromobacter cycloclastes*. *Science*, 253:438–442.

Gough, J. A. and Murray, N. E. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. *Journal of Molecular Biology*, 166:1–19.

Gow, A. J. and Stamler, J. S. (1998). Reactions between nitric oxide and haemoglobin under physiological conditions. *Nature*, 391:169–173.

Grigg, G. C., Wells, R. M. G., and Beard, L. A. (1993). Allosteric control of oxygen binding by haemoglobin during embryonic development in the crocodile *Crocodylus porosus*: the role of red cell organic phosphates and carbon dioxide. *Journal of Experimantal Biology*, 175:15–32.

Gros, P., van Gunsteren, W. F., and Hol, W. G. J. (1990). Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics. *Science*, 249:1149–1152.

Guo, D., Smith, G. D., Griffin, J. F., and Langs, D. A. (1995). Use of globic scattering factors for protein structures at low resolution. *Acta Crystallographica*, A51:945–947.

Halbrecht, I. and Klibanski, C. (1956). Identification of a new normal embryonic haemoglobin. *Nature*, 178(4537):794–795.

Harada, Y., Lifchitz, A., and Berthou, J. (1981). A translation function combining packing and diffraction information: an application to lysozyme (high-temperature form). *Acta Cryst.*, A37:398–406.

Haurowitz, F. (1938). Das gleichgewicht zwischen hamoglobin und sauerstoff. *Hoppe-Seyler's Z. Physiol. Chem.*, 254:266.

Hecht, F., Motulsky, A. G., Lemrie, R. J., and Shepard, T. E. (1966). Predominance of hemoglobin Gower 1 in early human embryonic development. *Science*, 152:91–92.

Heidner, F. H., Ladner, R. C., and Pertuz, M. F. (1976). Structure of horse carbonmonoxyhaemoglobin. *Journal of Molecular Biology*, 104:707–722.

Heilmeyer, L. (1943). *Spectrophotometry in Medicine*. Maclehose, Glasgow. Translated by A. Jordan and T. L. Tippell.

Heiss, B., Frunzke, K., and Zumft, W. G. (1989). Formation of the N-N bond from nitric oxide by a membrane-bound cytochrome *bc* complex of nitrate-respiring (denitrifying) *Pseudomonas stutzeri*. *Journal of Bacteriology*, 171:3288–3297.

Henry, Y. and Bessieres, P. (1984). Denitrification and nitrite reduction: *Pseudomonas aeruginosa* nitrite-reductase. *Biochimie*, 66:259–289.

Hill, A. V. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation current. *Journal of Physiology*, 40:4–7.

Hill, K. E. and Wharton, D. C. (1978). Reconstitution of the apoenzyme of cytochrome oxidase from *Pseudomonas aeruginosa* with heme $d_1$ and other heme groups. *The Journal of Biological Chemistry*, 253(2):489–495.

Hoffman, D. W., Cameron, C. S., Davies, C., White, S. W., and Ramakrishnan, V. (1996). Ribosomal protein L9: a structure determined by the combined use of X-ray crystallography and NMR spectroscopy. *Journal of Molecular Biology*, 264:1058–1071.

Hofmann, O. and Brittain, T. (1996). Ligand binding kinetics and dissociation of the human embryonic haemoglobins. *Biochemical Journal*, 315:65–70.

Hofmann, O., Carrucan, G., Robson, N., and Brittain, T. (1995a). The chloride effect in the human embryonic haemoglobins. *Biochemical Journal*, 309:959–962.

Hofmann, O., Mould, R., and Brittain, T. (1995b). Allosteric modulation of oxygen binding to the three human embryonic haemoglobins. *Biochemical Journal*, 306:367–370.

Hofmann, O. M., Mould, R. M., and Brittain, T. (1994). The incorporation of sulphaem into recombinant adult haemoglobin produced in a yeast expression system. *Protein Engineering*, 7(2):281–283.

Hole, U. H., Vollack, K-U., Zumft, W. G., Eisenmann, E., Siddiqui, R. A., Friedrich, B., and Kroneck, P. M. H. (1996). Characterization of the membranous denitrification enzymes nitrite reductase (cytochrome $cd_1$) and copper containing nitrous oxide reductase from *Thiobacillus denitrificans*. *Archives of Microbiology*, 165:55–61.

Hoppe-Seyler, F. (1864). Ueber die chemischen und optischen eigenschaften des blutfarbstoffs. *Virchow's Arch. Pathol. Anat.*, 29:233.

Horio, T., Higashi, T., , Sasagawa, M., Kusai, K., Nakai, M., and Okunuki, K. (1960). Purification and properties of cytochrome oxidase from *Pseudomonas aeruginosa*. *Biochemical Journal*, 77:194–201.

Huber, R. (1987). Flexibility and rigidity, requirements for the function of proteins and protein pigment complexes. *Biochemical Society Transactions*, 15:1009–1020.

Huehns, E. R., Dance, N., Beaven, G. H., Keil, J. V., Hecht, F., and Motulsky, A. G. (1964). Human embryonic haemoglobins. *Nature*, 201(4924):1095–1097.

Huehns, E. R. and Farooqui, A. M. (1975). Oxygen dissociation properties of human embryonic red cells. *Nature*, 254:335–337.

Huehns, E. R., Flynn, F. V., Butler, E. A., and Beaven, G. H. (1961). Two new haemoglobin variants in a very young human embryo. *Nature*, 189:496–497.

Imai, K. (1982). *Allosteric Effects in Haemoglobin*. Cambridge University Press.

Ingram, V. M. (1955). Sulphydryl groups in haemoglobins. *Biochemical Journal* , 59:653-661.

Ingram, V. M. (1956). A specific chemical difference between the globins of normal human and sickle cell anaemia haemoglobin. *Nature*, 178:792-794.

Ito, N., Phillips, S. E. V., Stevens, C., Ogel, Z. B., McPherson, M. J., Keen, J. N., Yadav, K. D. S., and Knowles, P. F. (1991). Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature*, 350:87–90.

Jancarik, J. and Kim, S-H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *Journal of Applied Crystallography*, 24:409–411.

Janin, J. and Wodak, S. J. (1993). The quaternary structure of carbonmonoxy hemoglobin Ypsilanti. *Proteins: Structure, Function and Genetics*, 15:1–4.

Jia, L., Bonaventura, C., Bonventura, J., and Stamler, J. S. (1996). S-nitrosohaemoglobin: a dynamic activity of blood involved in vascular control. *Nature*, 380:221–226.

Jiang, J-S. and Brunger, A. T. (1994). Protein hydration observed by X-ray diffraction. Solvation properties of penicillopepsin and neuraminidase crystal structures. *Journal of Molecular Biology*, 243:100–115.

Jones, S. and Thornton, J. M. (1995). Protein-protein interactions: A review of protein dimer structures. *Progress in Biophysics and Molecular Biology*, 63:31–65.

Jones, T. A. (1982). TOM: a graphics fitting program for macromolecules. In Sayre, D., editor, *Computational crystallography*, pages 303–317. Clarendon Press, Oxford.

Jungst, A., Braun, C., and Zumft, W. G. (1991a). Close linkage in *Pseudomonas stutzeri* of the structural genes for respiratory nitrite reductase and nitrous oxide reductase, and other essential genes for denitrification. *Molecular and General Genetics*, 225:241–248.

Jungst, A., Wakabayashi, S., Matsubara, H., and Zumft, W. G. (1991b). The *nirSTBM* region coding for cytochrome $cd_1$-dependent nitrite respiration of *Pseudomonas stutzeri* consists of a cluster of mono-, di-, and tetraheme proteins. *FEBS Letters*, 279:205–209.

Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica*, A32:922–923.

Kabsch, W. (1988). Automatic indexing of rotation diffraction patterns. *Journal of Applied Crystallography*, 21:67–71.

Kamuzora, H. and Lehmann, H. (1975). Human embryonic haemoglobins including a comparison by homology of the human ζ and α chains. *Nature*, 276:511–513.

Katz, D. S., White, S. P., Huang, W., Kumar, R., and Christianson, D. W. (1994). Structure determination of aquomet porcine hemoglobin at 2.8 Å resolution. *Journal of Molecular Biology*, 244:541–553.

Kavanaugh, J. S., Rogers, P. H., Case, D. A., and Arnone, A. (1992). High resolution X-ray study of deoxyhemoglobin Rothschild 37βTrp→Arg: a mutation that creates an intersubunit chloride-binding site. *Biochemistry*, 31:4111–4121.

Keilen, D. and Hartree, E. F. (1937). Reaction of nitric oxide with haemoglobin and methaemoglobin. *Nature*, 139:548.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. W., and Phillips, D. C. (1958). A three dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*, 181:662–666.

Kilmartin, J. V. and Rossi-Bernardi, L. (1969). Inhibition of $CO_2$ combination and reduction of the Bohr effect in haemoglobin chemically modified at its α-amino groups. *Nature*, 222:1243-1246.

Kingston, R. L. (1996). *X-Ray crystallographic investigations of the structures of enzymes of medical and biotechnological importance*. PhD thesis, Massey University, New Zealand.

Kingston, R. L., Baker, H. M., and Baker, E. N. (1994). Search designs for protein crystallization based on orthogonal arrays. *Acta Crystallographica*, D50:429–440.

Kingston, R. L., Scopes, R. K., and Baker, E. N. (1996). The structure of glucose fructose oxidoreductase from *Zymomonas mobilis*: an osmoprotective enzyme containing non-dissociable NADP. *Structure*, 4:1413–1428.

Kleywegt, G. J. (1996). Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallographica*, D52:842–857.

Kleywegt, G. J. (1997). Not your average density. *Structure*, 5:1557–1569.

Kleywegt, G. J. and Brunger, A. T. (1996). Checking your imagination: applications of the free R value. *Structure*, 4:897–904.

Kleywegt, G. J. and Jones, T. A. (1994). Halloween... masks and bones. In Bailey, S., Hubbard, R., and Waller, D., editors, *Proceedings of the CCP4 study weekend*, pages 59–66. EPSRC Daresbury Laboratory, Warrington, U.K.

Kleywegt, G. J. and Jones, T. A. (1995). Where freedom is given, liberties are taken. *Structure*, 3:535–540.

Kleywegt, G. J. and Jones, T. A. (1996). xdlMAPMAN and xdlDATAMAN - programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallographica*, D52:826–828.

Kleywegt, G. J. and Jones, T. A. (1997). Template convolution to enhance or detect structural features in macromolecular electron-density maps. *Acta Crystallographica*, D53:179–185.

Kleywegt, G. J., Zou, J-Y., Divne, C., Davies, G. J., Sinning, I., Stahlberg, J., Reinikainen, T., Srisodsuk, M., Teeri, T. T., and Jones, T. A. (1997). The crystal structure of the catalytic core domain of endoglucanase I from *Tricoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. *Journal of Molecular Biology*, 272:383–397.

Kobayashi, K., Koppenhofer, A., Ferguson, S. J., and Tagawa, S. (1997). Pulse radiolysis studies on cytochrome $cd_1$ nitrite reductase from *Thiosphaera pantotropha*: evidence for a fast intramolecular electron transfer from *c*-heme to $d_1$ heme. *Biochemistry*, 36:13611–13616.

Komer, H. and Zumft, W. (1989). Expression of denitrification enzymes in response to the dissolved oxygen level and respiratory substrate in continuous culture of *Pseudomonas stutzeri*. *Applied and Environmental Microbiology*, 55:1670–1676.

Kostrewa, D. (1997). Bulk solvent correction: practical application and effects in reciprocal and real space. In *CCP4 Newsletter No. 34 1997* pages 9–22. SERC Daresbury Laboratory, Warrington, U.K

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *Journal of Applied Crystallography*, 24:946–950.

Kuhn, E. P., Zeyer, J., Eicher, P., and Schwarzenbach, R. P. (1988). Anaerobic degradation of alkylated benzenes in denitrifying laboratory aquifer columns. *Applied and Environmental Microbiology*, 54:490–496.

Kukimoto, M., Nishiyama, M., Murphy, M. E. P., Turley, S., Adman, E. T., Horinouchi, S., and Beppu, T. (1994). X-ray structure and site directed mutagenesis of a nitrite reductase from *Alcaligenes faecalis* S-6: roles of two copper atoms in nitrite reduction. *Biochemistry*, 33:5246–5252.

Kunzer, W. (1957). Human embryo haemoglobins. *Nature*, 179:477–478.

Kuriyan, J., Osapay, K., Burley, S. K., Brunger, A. T., Hedrickson, W. A., and Karplus, M. (1991). Exploration of disorder in protein structures by X-ray restrained molecular dynamics. *Proteins: Structure, Function and Genetics*, 10:340–358.

Kuronen, T., Saraste, M., and Ellfolk, N. (1975). The subunit structure of *Pseudomonas* cytochrome oxidase.*Biochimica et Biophysica Acta*, 393:48–54.

Kuster, W. (1912). Beitrage zur kenntnis des bilirubins und hamins. *Hoppe-Seyler's Z. Physiol. Chem.*, 82:463.

Laskowski, R., MacArthur, M., Moss, D., & Thornton, J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26, 283–291.

Lehmann, H. and Huntsman, R. G. (1974). *Man's Haemoglobins*. North Holland Publishing Company.

Lemberg, R. and Falk, J. E. (1951). Comparison of haem *a*, the dichroic haem of heart muscle, and of porphyrin *a* with compounds of known structure. *Biochemical Journal*, 49:674–683.

Lesk, A. M. and Chothia, C. (1980). How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *Journal of Molecular Biology*, 136:225–270.

Li, J., Brick, P., O'Hare, M. C., Skarzynski, T., Lloyd, L. F., Curry, V. A., Clark, I. M., Bigg, H. F., Hazleman, B. L., Cawston, T. E., and Blow, D. M. (1995). Structure of full-length porcine synovial collagenase reveals a C-terminal domain containing a calcium-linked, four-bladed β-propeller. *Structure*, 3:341–349.

Liddington, R., Derewenda, Z., Dodson, E., Hubbard, R., and Dodson, G. (1992). High resolution crystal structures and comparisons of T-state deoxyhaemoglobin and two liganded T-state haemoglobins: T(α-oxy)haemoglobin and T(met)haemoglobin. *Journal of Molecular Biology*, 228:551–579.

Little, P. F. R., Flavell, R. A., Kooter, J. M., Annison, G., and Williamson, R. (1979). Structure of the human fetal globin gene locus. *Nature*, 278:227–231.

Liu, M. C., Bakel, B. W., Liu, M. Y., and Dao, T. N. (1988). Purification of *Vibrio fischeri* nitrite reductase and its characterization as a hexaheme *c*-type cytochrome. *Archives of Biochemistry and Biophysics*, 262:259–265.

Lu, G., Lindqvist, Y., Schneider, G., Dwivedi, U., and Campbell, W. (1995). Structural studies on corn nitrate reductase: refined structure of the cytochrome *b* reductase fragment at 2.5 Å, its ADP complex and an active-site mutant and modeling of the cytochrome *b* domain. *Journal of Molecular Biology*, 248:931–948.

Maclean, N. (1978). *Haemoglobin*. Edward Arnold Limited.

Makinen, M. W., Schichman, S. A., and Hill, S. C. (1983). Heme-heme orientation and electron transfer kinetic behavior of multisite oxidation-reduction enzymes. *Science*, 222:929–931.

Martinez, S. E., Huang, D., Szczepaniak, A., Cramer, W. A., and Smith, J. L. (1994). Crystal structure of chloroplast cytochrome *f* reveals a novel cytochrome fold and unexpected heme ligation. *Structure*, 2:95–105.

Mathews, F. S., Bethge, P. H., and Czerwinski, E. W. (1979). The structure of cytochrome $b_{562}$ from *Escherichia coli* at 2.5 Å resolution. *The Journal of Biological Chemistry*, 254:1699–1706.

Mathews, F. S., Chen, Z. W., Bellamy, H. D., and McIntire, W. S. (1991). Three-dimensional structure of p-cresol methylhydroxylase (flavocytochrome *c*) from *Pseudomonas putida* at 3.0-Å resolution. *Biochemistry*, 30:238–247.

Matsuura, Y., Takano, T., and Dickerson, R. E. (1982). Structure of cytochrome *c*-551 from *P. aeruginosa* refined at 1.6 Å resolution and comparison of the two redox forms. *Journal of Molecular Biology*, 156:389–409.

Matthews, B. (1968). Solvent content of protein crystals. *The Journal of Molecular Biology*, 33:491–497.

Melderis, H., Steinheider, G., and Ostertag, W. (1974). Evidence for a unique kind of α-type globin chain in early mammalian embryos. *Nature*, 250:774–776.

Meyer, T. E. and Kamen, M. D. (1982). New perspectives on *c*-type cytochromes. *Advances in Protein Chemistry*, 35:105–212.

Moews, P. C. and Kretsinger, R. H. (1975). Refinement of the structure of carp muscle calcium-binding parvalbumin by model building and difference Fourier analysis. *Journal of Molecular Biology*, 91:201–228.

Moore, G. R. and Pettigrew, G. W. (1990). *Cytochromes c: evolutionary, structural and physicochemical aspects*. Springer-Verlag.

Moser, C. C., Keske, J. M., Warncke, K., Farid, R. S., and Dutton, P. L. (1992). Nature of biological electron transfer. *Nature*, 355:796–802.

Mould, R. M., Hofmann, O. M., and Brittain, T. (1994). Production of human embryonic haemoglobin (Gower II) in a yeast expression system. *Biochemical Journal*, 298:619–622.

Muirhead, H. and Perutz, M. F. (1963). Structure of hemoglobin: a three dimensional Fourier synthesis of reduced human haemoglobin at 5.5 Å resolution. *Nature*, 199(4894):633–639.

Murshudov, G. N., Dodson, E. J., and Vagin, A. A. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica*, D53:240–255.

Navaza, J. (1994). AMORE: an automated package for molecular replacement. *Acta Crystallographica*, A50:157–163.

Neer, J. N. and Smith, T. F. (1996). G protein heterodimers: new structures propel new questions. *Cell*, 84:175–178.

Nicholls, A., Sharp, K. A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Structure, Function, and Genetics*, 11:281–296.

Nurizzo, D., Silvestrini, M. C., Mathieu, M., Cutruzzola, F., Bourgeois, D., Fulop, V., Hajdu, J., Brunori, M., Tegoni, M., and Cambillau, C. (1997). Crystal structure of oxidised nitrite reductase from *Pseudomonas aeruginosa* at 2.15 Å resolution. *Structure*, 5(9):1157–1171.

Ostermeier, C., Harrenga, A., Ermler, U., and Michel, H. (1997). Structure at 2.7 Å resolution of the *Paracoccus denitrificans* two-subunit cytochrome *c* oxidase complexed with an antibody Fv fragment. *Proceedings of the National Academy of Sciences, USA*, 94:10547–10553.

Otwinowski, Z. (1993). Oscillation data reduction program. In Sawyer, L., Isaacs, N., and Bailey, S., editors, *Proceedings of the CCP4 Study Weekend*, pages 56–62. SERC Daresbury Laboratory, Warrington, U.K.

Palmer, R. M., Ferrige, A. G., and Moncada, S. (1987). Nitric oxide release accounts for the biological activity of endothelium-derived relaxing factor. *Nature*, 327:524–526.

Pannu, N. S. and Read, R. J. (1996). Improved structure refinement through maximum likelihood. *Acta Crystallographica*, A52:659–668.

Paoli, M., Liddington, R., Tame, J., Wilkinson, A., and Dodson, G. (1996). Crystal structure of T state haemoglobin with oxygen bound at all four haems. *Journal of Molecular Biology*, 256:775–792.

Parr, S. R., Barber, D., and Greenwood, C. (1976). A purification procedure for the soluble cytochrome oxidase and some other respiratory proteins from *Pseudomonas aeruginosa*. *Biochemical Journal*, 157:423–430.

Parr, S. R., Barber, D., Greenwood, C., and Brunori, M. (1977). The electron-transfer reaction between azurin and the cytochrome *c* oxidase from *Pseudomonas aeruginosa*. *Biochemical Journal*, 167:447–455.

Parr, S. R., Wilson, M. T., and Greenwood, C. (1975). The reaction of *Pseudomonas aeruginosa* cytochrome *c* oxidase with carbon monoxide. *Biochemical Journal*, 151:51–59.

Pelletier, H. and Kraut, J. (1992). Crystal structure of a complex between electron transfer partners, cytochrome *c* peroxidase and cytochrome *c*. *Science*, 258:1748–1755.

Perrella, M., Benazzi, L., Ripamonti, M., and Rossi-Bernardi, L. (1994). Bohr effect in hemoglobin deoxy/cyanomet intermediates. *Biochemistry*, 33:10358–10366.

Perutz, M. F., Fermi, G., Luisi, B., Shaanan, B., and Liddington, R. C. (1987). Stereochemistry of cooperative mechanisms in haemoglobin. *Accounts of Chemical Research*, 20(9):309–321.

Perutz, M. F. (1968). Preparation of haemoglobin crystals. *Journal of Crystal Growth*, 2:54–56.

Perutz, M. F. (1970). The Bohr effect and combination with organic phosphates. *Nature*, 228:734–739.

Perutz, M. F. (1972). Nature of haem-haem interaction. *Nature*, 237:495–499.

Perutz, M. F. (1979). Regulation of oxygen affinity of haemoglobin: influence of structure of the globin on the heme iron. *Annual Reviews in Biochemistry*, 48:327–386.

Perutz, M. F. (1996). Blood: taking the pressure off. *Nature*, 380:205–206.

Perutz, M. F., Fermi, G., Poyart, C., Pagnier, J., and Kister, J. (1993). A novel allosteric mechanism in haemoglobin. Structure of bovine deoxyhaemoglobin, absence of specific chloride-binding sites and origin of the chloride-linked Bohr effect in bovine and human haemoglobin. *Journal of Molecular Biology*, 233:536–545.

Perutz, M. F., Kilmartin, J., Nishikura, K., Fogg, J., and Butler, P. J. G. (1980). Identification of residues contributing to the Bohr effect of human haemoglobin. *Journal of Molecular Biology*, 138:649–670.

Perutz, M. F., Muirhead, H., Mazzarella, L., Crowther, R. A., Greer, J., and Kilmartin, J. V. (1969). Identification of residues responsible for the alkaline Bohr effect in haemoglobin. *Nature*, 222:1240–1243.

Perutz, M. F., Rossman, M. G., Cullis, A., Muirhead, H., and Will, G. (1960). Structure of haemoglobin: a three dimensional Fourier synthesis at 5.5 Å resolution, obtained by X-ray analysis. *Nature*, 185:416–422.

Perutz, M. F., Shih, D. T., and Williamson, D. (1994). The chloride effect in human haemoglobin. A new kind of allosteric mechanism. *Journal of Molecular Biology*, 239:555–560.

Peschle, C., Mavilio, F., Care, A., Migliacco, G., Migliaccio, A. R., Salvo, G., Samoggia, P., Petti, S., Guerrio, R., Marinucci, M., Lazzaro, D., Russo, G., and Mastroberardino, G. (1985). Haemoglobin switching in human embryos: asynchrony of $\zeta \rightarrow \alpha$ and $\varepsilon \rightarrow \gamma$-globin switches in primitive and definitive erythropoietic lineage. *Nature*, 313:235–238.

Peschle, C., Migliaccio, A. R., Migliaccio, G., Petrini, M., Calandrini, M., Russo, G., Mastroberardino, G., Presta, M., Gianni, A. M., Comi, P., Giglioni, B., and Ottolenghi, S. (1984). Embryonic → fetal Hb switch in humans: studies on erythroid bursts generated by embryonic progenitors from yolk sac and liver. *Proceedings of the National Academy of Science, USA*, 81:2416–2420.

Pettigrew, G. W. and Moore, G. R. (1987). *Cytochromes c: biological aspects*. Springer-Verlag.

Poulos, T. L. (1988). Heme enzyme crystal structures. In Eichorn, G. L. and Marzilli, L. G., editors, *Heme Proteins*. Elsevier Science Publishing.

Poulos, T. L., Sheriff, S., and Howard, A. J. (1987). Cocrystals of yeast cytochrome *c* peroxidase and horse heart cytochrome *c*. *The Journal of Biological Chemistry*, 262:13881–13884.

Proudfoot, N. J., Gil, A., and Maniatis, T. (1982). The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. *Cell*, 31:553–563.

Ravichandran, K. G., Boddupalli, S. S., Hasemann, C. A., Peterson, J. A., and Deisenhofer, J. (1993). Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science*, 261:731–736.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallographica*, A42:140–149.

Read, R. J. (1990). Structure-factor probabilities for related structures. *Acta Crystallographica*, A46:900–912.

Rees, D. C., Lewis, M., and Lipscomb, W. N. (1983). Refined crystal structure of carboxypeptidase A at 1.54 Å resolution.*Journal of Molecular Biology*, 168:367–387.

Rees, E., Siddiqui, R. A., Koster, F., Schneider, B., and Friedrich, B. (1997). Structural gene (*nirS*) for the cytochrome $cd_1$ nitrite reductase of *Alcaligenes eutrophus*. *Applied and Environmental Microbiology*, 63(2):800–802.

Reichert, E. T. and Brown, A. P. (1909). *The Differentiation and Specificity of Corresponding Proteins and Other Vital Substances in Relation to Biological Classification and Organic Evolution: The Crystallography of Hemoglobins*. Carnegie Institute of Washington Publishing.

Renault, L., Nassar, N., Vetter, I., Becker, J., Klebe, C., Roth, M., and Wittinghofer, A. (1998). The 1.7 Å crystal structure of the regulator of chromosome condensation (RCC1) reveals a seven- bladed propeller. *Nature*, 392:97–100.

Rhinesmith, H. S., Schroeder, W. A., and Pauling, L. (1957). A quantitative study of the hydrolysis of human dinitrophenyl (DNP) globin: the number and kind of polypeptide chains in normal adult human hemoglobin. *Journal of the American Chemical Society*, 79:4682.

Rice, L. M. and Brunger, A. T. (1994). Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins*, 19:277–290.

Rifkind, J. M. (1988). Hemoglobin. In Eichorn, G. L. and Marzilli, L. G., editors, *Heme Proteins*. Elsevier Science Publishing.

Robertson, L. A. and Kuenen, J. G. (1984). Aerobic denitrification: a controversy revived. *Archives of Microbiology*, 139:351–354.

Robson, N. and Brittain, T. (1996). Heme stability in the human embryonic hemoglobins. *Journal of Inorganic Biochemistry*, 64:137–147.

Roggentin, P., Rothe, B., Kaper, J. B., Galen, J., Lawrisuk, L., Vimr, E. R., and Schauer, R. (1989). Conserved sequences in bacterial and viral sialidases. *Glyconconjugate Journal*, 6:349–353.

Rollema, H. S., de Bruin, S. H., Janssen, L. H., and van Os., G. A. (1975). The effect of potassium chloride on the Bohr effect of human hemoglobin. *The Journal of Biological Chemistry*, 250:1333–1339.

Rossmann, M. G. and Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica*, 15:24–31.

Rost, B. and Sander, C. (1994a). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins: Structure, Function and Genetics*, 19:55–72.

Rost, B. and Sander, C. (1994b). Conservation and prediction of solvent accessibility in protein families. *Proteins: Structure, Function and Genetics*, 20:216–226.

Roussel, A. and Cambillau, C. (1991). *Silicon Graphics Geometry Partners Directory*, pages 77–78. Silicon Graphics, Mountain View, CA.

Rudenko, G., Bonten, E., D'Azzo, A., and Hol, W. G. J. (1996). Structure determinination of the human protective protein: twofold averaging reveals the three-dimensional structure of a domain which was entirely absent in the initial model. *Acta Crystallographica*, D52:923–936.

Sakabe, N. (1991). X-ray diffraction data collection system for modern protein crystallography with a Weissenberg camera and an imaging plate using synchrotron radiation. *Nuclear Instruments and Methods in Physics Research*, A303:448–463.

Saraste, M., Virtanen, I., and Kuronen, T. (1977). The quaternary structure of *Pseudomonas* cytochrome oxidase studied by electron microscopy. *Biochimica et Biophysica acta*, 492:156–162.

Sato, M., Yamamoto, M., Imada, K., Katsube, Y., Tanaka, N., and Higashi, T. (1992). A high speed data-collection system for large-unit-cell crystals using an imaging plate as a detector. *Journal of Applied Crystallography*, 25:348–357.

Scheepens, A., Mould, R., Hofmann, O., and Brittain, T. (1995). Some effects of post-translational N-terminal acetylation of the human embryonic ζ globin protein. *Biochemical Journal*, 310:597-600.

Schichman, S. A., Meyer, T. E., and Gray, H. B. (1996). Kinetics of electron transfer in *Pseudomonas aeruginosa* cytochrome $cd_1$-nitrite reductase. *Inorganica Chimica Acta*, 243:25–31.

Schindelin, H., Kisker, C., Hilton, J., Rajagopalan, K. V., and Rees, D. C. (1996). Crystal structure of DMSO reductase: redox-linked changes in molybdopterin coordination. *Science*, 272:1615–1621.

Schroeder, W. A., Shelton, J. R., Shelton, J. B., Cormick, J., and Jones, R. T. (1963). The amino acid sequence of the γ chain of human fetal hemoglobin. *Biochemistry*, 2(5):992–1008.

Schumacher, M. A., Zheleznova, E. E., Poundstone, K. S., Kluger, R., Jones, R. T., and Brennan, R. G. (1997). Allosteric intermediates indicate R2 is the liganded hemoglobin end state. *Proceedings of The National Academy of Sciences, USA*, 94:7841–7844.

Shaanan, B. (1983). Structure of human oxyhaemoglobin at 2.1 Å resolution. *Journal of Molecular Biology*, 171:31–61.

Sheldrick, G M. (1997). SHELX-97. University of Gottingen.

Sheriff, S. and Hendrickson, W. A. (1987). Description of overall anisotropy in diffraction from macromolecular crystals. *Acta crystallographica*, A43:118–121.

Shih, D. T. and Perutz, M. F. (1987). Influence of anions and protons on the Adair coefficients of haemoglobins A and Cowtown (His HC3(146)β → Leu). *Journal of Molecular Biology*, 195:419–422.

Shimada, H. and Orii, Y. (1975). The nitric oxide compounds of *Pseudomonas aeruginosa* nitrite reductase and their participation in nitrite reduction. *FEBS letters*, 54:237–240.

Silva, M. M., Rogers, P. H., and Arnone, A. (1992). A third quaternary structure of human hemoglobin A at 1.7 Å resolution. *The Journal of Biological Chemistry*, 267(24):17248–17256.

Silvestrini, M. C., Colosimo, A., Brunori, M., Walsh, T. A., Barber, D., and Greenwood, C. (1979). A reevaluation of some basic structural and functional properties of *Pseudomonas* cytochrome oxidase. *Biochemical Journal*, 183:701–709.

Silvestrini, M. C., Cutruzzola, F., Schinina, M. E., Maras, B., Rolli, G., and Brunori, M. (1996). Isolation and characterization of the $d_1$ domain of *Pseudomonas aeruginosa* nitrite reductase. *Journal of Inorganic Biochemistry*, 62:77–87.

Silvestrini, M. C., Falcinelli, S., Ciabatti, I., Cutruzzola, F., and Brunori, M. (1994). *Pseudomonas aeruginosa* nitrite reductase (or cytochrome oxidase): an overview. *Biochimie*, 76:641–654.

Silvestrini, M. C., Galeotti, C.L., Gervais, M., Schinina, E., Barra, D., Bossa, F., and Brunori, M. (1989). Nitrite reductase from *Pseudomonas aeruginosa*: sequence of the gene and the protein. *FEBS Letters*, 254:33–38.

Silvestrini, M. C., Tegoni, M., Celerier, J., Desbois, A., and Gervais, M. (1993). Expression in *Escherichia coli* of the flavin and the haem domains of *Hansenula anomala* flavocytochrome $b_2$ (flavodehydrogenase and $b_2$ core) and characterization of the recombinant proteins. *Biochemical Journal*, 293:501–508.

Silvestrini, M. C., Tordi, M. G., Citro, G., Vecchini, P., and Brunori, M. (1995). Monomeric *Pseudomonas aeruginosa* nitrite reductase: preparation, characterization, and kinetic properties. *Journal of Inorganic Biochemistry*, 57:169–181.

Silvestrini, M. C., Tordi, M. G., Colosimo, A., Antonini, E., and Brunori, M. (1982). The kinetics of electron transfer between *Pseudomonas aeruginosa* cytochrome *c*-551 and its oxidase. *Biochemical Journal*, 203:445–451.

Silvestrini, M. C., Tordi, M. G., Musci, G., and Brunori, M. (1990). The reaction of *Pseudomonas* nitrite reductase and nitrite, a stopped-flow and EPR study. *The Journal of Biological Chemistry*, 265(20):11783–11787.

Smith, F. R., Lattman, E. E., and Carter, C. W. (1991). The mutation β99 Asp-Tyr stabilizes Y, a new composite quaternary state of human hemoglobin. *Proteins: Structure, Function and Genetics*, 10:81–91.

Smith, F. R. and Simmons, K. C. (1994). Cyanomet human hemoglobin crystallized under physiological conditions exhibits the Y quaternary state. *Proteins: Structure, Function and Genetics*, 18:295–300.

Smith, G. B. and Tiedje, J. M. (1992). Isolation and characterization of a nitrite reductase gene and its use as a probe for denitrifying bacteria. *Applied and Environmental Microbiology*, 58(1):376–384.

Sousa, R. (1995). Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization. *Acta Crystallographica*, D51:271–277.

Springer, T. A. (1997). Folding of the N-terminal, ligand-binding region of integrin α-subunits into a β-propeller domain. *Proceedings of the National Academy of Science, USA*, 94:65–72.

Srinivasan, R. and Rose, G. D. (1994). The T-to-R transformation in hemoglobin: a reevaluation. *Proceedings of the National Academy of Sciences, USA*, 91:11113–11117.

Stamatoyannopoulos, G., Constantoulakis, P., Brice, M., Kurachi, S., and Papayannopoulos, T. (1987). Coexpression of embryonic, fetal, and adult globins in erythroid cell of human embryos: relevance to the cell-lineage models of globin switching. *Developmental Biology*, 123:191–197.

Stamatoyannopoulos, G., Rosenblum, B. B., Papayannopoulos, T. H., Brice, M., Nakamoto, B., and Shepard, T. H. (1979). Hb F and Hb A production in erthyroid culture from human fetuses and neonates. *Blood*, 54:440–449.

Stamler, J. S., Jia, L., Eu, J. P., McMahon, T. J., Demchenko, I. T., Bonaventura, J., Gernert, K., and Piantadosi, C. A. (1997). Blood flow regulation by S-nitrosohemoglobin in the physiological oxygen gradient. *Science*, 276:2034–2037.

Stokes, G. G. (1864). On the reduction and oxygenation of the colouring matter of the blood. *Proceedings of the Royal Society, London*, 13:355.

Stryer, L. (1995). *Biochemistry*. W. H. Freeman and Company, New York, 4th edition.

Sutherland, J., Greenwood, C., Peterson, J., and Thompson, A. J. (1986). An investigation of the ligand-binding properties of *Pseudomonas aeruginosa* nitrite reductase. *Biochemical Journal*, 233:893–898.

Svedberg, T. and Fahraeus, R. (1926). A new method for the determination of the molecular weight of the proteins. *Journal of the American Chemical Society*., 48:430.

Takano, T. and Dickerson, R. E. (1981). Conformation change of cytochrome *c*: ferrocytochrome *c* structure refined at 1.5 Å resolution. *Journal of Molecular Biology*, 153:79–94.

Tame, J. R. H., Wilson, J. C., and Weber, R. E. (1996). The crystal structures of trout Hb I in the deoxy and carbonmonoxy forms. *Journal of Molecular Biology*, 259:749–760.

Tegoni, M. and Cambillau, C. (1994). The 2.6-Å refined structure of the *Escherichia coli* recombinant *Saccharomyces cerevisiae* flavocytochrome $b_2$-sulfite complex. *Protein Science*, 3:303–313.

Teichman, N. L. (1853). Ueber die crystallisation der organischen bestandtheile des bluts. *Z. Ration. Med.*, 3:374.

Ten Eyck, L. F. (1995). Gro-ing pains. *Nature Structural Biology*, 2:1038–1042.

Tesmer, J. J., Sunahara, R. K., Gilman, A. G., and Sprang, S. R. (1997). Crystal structure of the catalytic domains of adenylyl cyclase in a complex with $G_{s\alpha}$.GTP$\gamma$S. *Science*, 278:1907–1916.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22:4673–4680.

Timkovich, R. and Cai, M. (1993). Investigation of the structure of oxidized *Pseudomonas aeruginosa* cytochrome *c*-551 by NMR: comparison of observed paramagnetic shifts and calculated pseudocontact shifts. *Biochemistry*, 32:11516–11523.

Tollin, P. and Rossmann, M. G. (1966). A description of various rotation function programs. *Acta Crystallographica*, 21:872–876.

Tong, L. and Rossmann, M. G. (1990). The locked rotation function. *Acta Crystallographica*, A46:783–792.

Tordi, M. G., Silvestrini, M. C., Colosimo, A., Tuttobello, L., and Brunori, M. (1985). Cytochrome *c*-551 and azurin oxidation catalysed by *Pseudomonas aeruginosa* cytochrome oxidase. *Biochemical Journal*, 230:797–805.

Tronrud, D. E. (1996a). Knowledge-based B-factor restraints for the refinement of proteins. *Journal of Applied Crystallography*, 29:100–104.

Tronrud, D. E. (1996b). The limits of interpretation. In Dodson, E., Moore, M., Ralph, A., and Bailey, S., editors, *Proceedings of the CCP4 Study Weekend*, pages 1–10. CCLRC Daresbury Laboratory, Warrington, U.K.

Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987). An efficient general-purpose least squares refinement program for macromolecular structures. *Acta Crystallographica*, A43:489–501.

Tschech, A. and Fuchs, G. (1987). Anaerobic degradation of phenol by pure cultures of newly isolated denitrifying pseudomonads. *Archives of Microbiology*, 148:213–217.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., and Yoshikawa, S. (1996). The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 Å. *Science*, 272:1136–1137.

Tuchinda, S., Nagai, K., and Lehmann, H. (1975). Oxygen dissociation curve of haemoglobin Portland. *FEBS Letters*, 49(3):390–391.

Tyuma, I. and Shimizu, K. (1970). Effect of organic phosphates on the difference in oxygen affinity between fetal and adult human hemoglobin. *Federation Proceedings*, 29:1112–1114.

Urzhumtsev, A. G. and Podjamy, A. D. (1996). On the problem of solvent modelling in macromolecular crystals using diffraction data: 1. The low resolution range. *CCP4 Newsletter on protein crystallography*, 31:12–16.

Vainshtein, B. K., Melik-Adamyan, W. R., Barynin, V. V., Vagin, A. A., Grebenko, A. I., and Borisov, V. V. (1986). Three-dimensional structure of catalase from *Penicillium vitale* at 2.0 Å resolution. *Journal of Molecular Biology*, 188:49–61.

Valegard, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J., and Liljas, L. (1994). Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature*, 371:858–863.

Van de Kamp, M., Silvestrini, M. C., Brunori, M., Van Beeumen, J., Hali, F. C., and Canters, G. W. (1990). Involvement of the hydrophobic patch of azurin in the electron-transfer reactions with $c_{551}$ and nitrite reductase. *European Journal of Biochemistry*, 194:109–118.

Varghese, J. N., Laver, W. G., and Colman P. M. (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature*, 303:35–40.

Vellieux, F. M. D., Huitema, F., Groendijk, H., Kalk, K. H., Frank, J. J., Jongejan, J. A., Duine, J. A., Petratos, K., Drenth, J., and Hol, W. G. J. (1989). Structure of a quinoprotein methylamine dehydrogenase at 2.25 Å resolution. *EMBO Journal*, 8:2171–2178.

Wagenbach, M., O'Rourke, K., Vitez, L., Wieczorek, A., Hoffman, S., Durfee, S., Tedesco, J., and Stetler, G. (1991). Synthesis of wild type and mutant human hemoglobins in *Saccharomyces cerevisiae*. *Bio/Technology*, 9:57–61.

Walker, G. and Nicholas, D. J. D. (1960). Nitrite reductase from *Pseudomonas aeruginosa*. *Biochemical Journal*, 77:4p-5p.

Wall, M. A., Coleman, D. E., Lee, E., Iniquez-Lluhi, J. A., Posner, B. A., Gilman, A. G., and Sprang, S. R. (1995). The structure of the G protein heterotrimer $G\alpha i_1\beta_1\gamma_2$. *Cell*, 83:1047–1058.

Wall, M. E., Clarage, J. B., and Phillips, G. N. (1997). Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering. *Structure*, 5(12):1599–1612.

Walsh, T. A., Johnson, M. K., Barber, D., Thomson, A. J., and Greenwood, C. (1980). Studies on heme $d_1$ extracted from *Pseudomonas aeruginosa* nitrite reductase. *Journal of Inorganic Chemistry*, 14:15–31.

Walsh, T. A., Johnson, M. K., Greenwood, C., Barber, D., Springall, J. P., and Thomson, A. J. (1979). Some magnetic properties of *Pseudomonas cytochrome* oxidase. *Biochemical Journal*, 177:29–39.

Wang, X. and Janin, J. (1993). Orientation of non-crystallographic symmetry axes in protein crystals. *Acta Crystallographica*, D49:505–512.

Wanner, U., Kemmler, J., Weilenmann, H. U., Egli, T., El-Banna, T., and Auling, G. (1990). Isolation and growth of a bacterium able to degrade nitriloacetic acid under denitrifying conditions. *Biodegradation*, 1:31–41.

Weatherall, D. J., editor (1976). *Haemoglobin: Structure, Function and Synthesis*, volume 32. The Medical Department, The British Council.

Weber, P. C., Bartsch, R. G., Cusanovich, M. A., Hamlin, R. C., Howard, A., Jordan, S. R., Kamen, M. D., Meyer, T. E., Weatherford, D. W., and Salemme, F. R. (1980). Structure of cytochrome *c'*: a dimeric, high-spin haem protein. *Nature*, 286:302–304.

Weeg-Aerssens, E., Wu, W., Ye, R. W., Tiedje, J. M., and Chang, C. K. (1991). Purification of cytochrome $cd_1$ nitrite reductase from *Pseudomonas stutzeri* JM300 and reconstitution with native and synthetic heme $d_1$. *The Journal of Biological Chemistry*, 266(12):7496–7502.

Williams, P. A., Fulop, V., Garman, E. F., Saunders, N. F. W., Ferguson, S. J., and Hajdu, J. (1997). Haem-ligand switching during catalysis in crystals of a nitrogen-cycle enzyme. *Nature*, 389:406–412.

Williams, P. A., Fulop, V., Leung, Y-C., Chan, C., Moir, J. W. B., Howlett, G., Ferguson, S. J., Radford, S. E., and Hajdu, J. (1995). Pseudospecific docking surfaces on electron transfer proteins as illustrated by pseudoazurin, cytochrome $c_{550}$ and cytochrome $cd_1$ nitrite reductase. *Nature Structural Biology*, 2(11):975–982.

Wilson, A. J. C. (1942). Determination of absolute from relative X-ray intensity data. *Nature*, 150:152.

Winterhalter, K. H., Ioppolo, C., and Antonini, E. (1971). Distribution of heme in systems containing heme-free and heme-bound hemoglobin chains. *Biochemistry*, 10:3790–3795.

Wood, W. G., Bunch, C., Kelly, S., Gunn, Y., and Breckon, G. (1985). Control of haemoglobin switching by a developmental clock. *Nature*, 313:320–323.

Xia, Z., Dai, W., Xiong, J., Hao, Z., White, V. L., and Mathews, F. S. (1992). The three dimensional structures of methanol dehydrogenase from two methlotrophic bacteria at 2.6 Å resolution. *The Journal of Biological Chemistry*, 267:22289–22297.

Xia, Z. and Mathews, F. S. (1990). Molecular structure of flavocytochrome $b_2$ at 2.4 Å resolution. *Journal of Molecular Biology*, 212:837–863.

Yamanaka, T., Ota, A., and Okunuki, K. (1961). A nitrite reducing system reconstructed with purified cytochrome components of *Pseudomonas aeruginosa*. *Biochim. Biophys. Acta*, 53:294–308.

Ye, R. W., Averill, B. A., and Tiedje, J. M. (1994). Denitrification: production and consumption of nitric oxide. *Applied and Environmental Microbiology*, 60(4):1053–1058.

Zhang, X.J. and Matthews, B. W. (1994). Enhancement of the method of molecular replacement by incorporation of known structural information. *Acta Crystallographica*, D50:675–686.

Zumft, W. G., Braun, C., and Cuypers, H. (1994). Nitric oxide reductase from *Pseudomonas stutzeri*. Primary structure and gene organisation of a novel bacterial cytochrome *bc* complex. *European Journal of Biochemistry*, 219:481–490.